Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AI/ML QSAR Modeling and Virtual Screening of p38α Kinase (MAPK14) Inhibitors Using a Leakage-Free 100-Model Ensemble

K. R. Jinuraj[1]
Computational Chemistry & AI Drug Discovery Laboratory
Shashidhar N. Rao[2]
snrao@ucsd.edu

University of California San Diego
Vitukudi N. Balaji[1]
Computational Chemistry & AI Drug Discovery Laboratory

Research Article

Keywords: Mitogen-activated protein kinase 14, MAPK14, CHEMBL260, 100-Model QSAR Benchmark, Tanimoto Similarity Kernels, Activity-Cliff-Tolerant GBDT, FT-Transformer, Stacking Ensemble

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11748308/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

AI/ML QSAR Modeling and Virtual Screening of p38α Kinase (MAPK14) Inhibitors Using a Leakage-Free 100-Model Ensemble

K. R. Jinuraj[1] · Shashidhar N. Rao[2] · Vitukudi N. Balaji[1]*

[1] Computational Chemistry & AI Drug Discovery Laboratory, No. 579, First Main Road, II Stage, III Block, RMV Extension, Bengaluru 560094, Karnataka, India

[2] Department of Bioengineering, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

* Corresponding authors: Vitukudi N. Balaji (e-mail: vnbalaji@gmail.com); Shashidhar N. Rao (e-mail: snrao@ucsd.edu)

Keywords Mitogen-activated protein kinase 14 · MAPK14 · CHEMBL260 · 100-Model QSAR Benchmark · Tanimoto Similarity Kernels · Activity-Cliff-Tolerant GBDT · FT-Transformer · Stacking Ensemble

Abstract

Mitogen-activated protein kinase 14 (MAPK14, ChEMBL ID: CHEMBL260) is a serine-threonine kinase whose dysregulated activation underpins chronic inflammation, neurodegeneration, and chemoresistance, yet the absence of selective small-molecule inhibitors has constrained translational progress. Here we present a standalone, 100-architecture quantitative structure–activity relationship (QSAR) and virtual screening study spanning 10 algorithmic families and 7 feature regimes, constructed exclusively from 3,504 curated ChEMBL bioactivity compounds partitioned into 2,804 training and 700 validation entries. A leakage-free, five-stage feature-selection funnel reduced 4,313 raw descriptors to seven non-redundant feature regimes, each independently propagated through the full architectural battery. The 100 models encompass Tanimoto and MinMax similarity kernels, cliff-tolerant Huber- and MAE-regularised gradient-boosted decision trees with DART stochasticity, an FT-Transformer attention backbone, a Dual-Stream Cross-Attention network, and a Grand 10-Family Stacking meta-learner. The Top-5 Consolidated Ensemble achieved 94.0% Top-1% active recovery and 94.0% Union recovery on the held-out validation set. Two independent evaluation events were conducted: (i) a retrospective decoy-enrichment assay in which 50 known active inhibitors were seeded among 4,950 diverse background decoys (total pool = 5,000) to benchmark the ensemble's ranking fidelity, and (ii) a prospective virtual screen of the KLIFS database comprising 2,379 unique kinase inhibitors to identify previously unreported MAPK14 hits. The results establish a precise, reproducible computational framework for selective MAPK14 inhibitor prioritisation and provide a rigorous benchmark against which future structure-based and ligand-based campaigns may be calibrated.

1. Introduction

Mitogen-activated protein kinase 14 (MAPK14, ChEMBL ID: CHEMBL260), widely recognised in the primary literature under the synonyms CSBP, CSBP1, CSBP2, CSPB1, EXIP, MAP kinase 14, MAP

kinase MXI2, MAPK 14, MAPK14, Mxi2, p38, p38 MAP kinase, p38a, p38α, and p38α Exip, is a 38-kDa serine-threonine protein kinase that occupies a central node in the tumour necrosis factor receptor-associated factor 6 (TRAF6) and tumour necrosis factor receptor-associated factor 2 (TRAF2)–mediated stress-kinase cascade [1–3]. Upon dual phosphorylation at the activation loop, MAPK14 transduces genotoxic, osmotic, and inflammatory stimuli to downstream effectors including the mitogen-activated protein kinase-activated protein kinase 2 (MK2) and the transcription factor ATF2, thereby governing nuclear F-actin remodelling during DNA-damage response [1], epithelial–mesenchymal transition in antimony-induced pulmonary fibrosis [3], and the self-amplifying Col23a1–DDR1–p38/ERK positive-feedback loop that sustains alveolar epithelial remodelling [3]. In the oncological setting, the p38α MAPK–MK2 axis has been implicated in platinum-chemotherapy-mediated nuclear actin assembly that confers ovarian cancer chemoresistance [1], while in hepatocellular carcinoma the cyclic lipopeptide mycosubtilin engages ER-stress-associated MAPK signalling to suppress HepG2 proliferation [4]. In neurodegenerative and autoimmune contexts, MAPK14 activity contributes to τ hyperphosphorylation and neurofibrillary tangle formation in Alzheimer's disease [5, 6], to synovial fibroblast-like synoviocyte apoptosis in rheumatoid arthritis [7], and to macrophage NF-κB and AKT1 cross-talk in experimental periodontitis [8, 9]. Collectively, these precise mechanistic observations underscore MAPK14 as a validated therapeutic target whose selective pharmacological modulation remains an unmet clinical need.

Prior computational campaigns targeting MAPK14 have employed a heterogeneous array of structure-based and ligand-based strategies, each with distinct scope and limitations. Vermani et al. [10] applied a deep neural network trained on fingerprint-clustered bioactivity data to repurpose FDA-approved drugs, identifying 149 candidate p38 inhibitors whose docking scores exceeded those of random FDA compounds. Badrinarayan and Sastry [11] developed a three-step scoring-and-filtering protocol for type II (DFG-out) p38 MAPK inhibitors, extracting 98 allosteric fragments from a 10^7-compound virtual screen and merging them with ATP-site binders via a urea linker. Pinsetta et al. [6] combined structure- and ligand-based design with molecular dynamics and molecular interaction field analyses to select thirteen low-toxicity candidates for Alzheimer's disease applications. Cheeseright et al. [12] applied the FieldScreen molecular-field-similarity method to 1.2 million compounds, yielding a thiadiazole lead with pIC_{50} of 6.4. Choi et al. [13] screened approximately 240,000 compounds with a modified AUTODOCK scoring function, recovering six KINOMEscan-confirmed hits with IC50 values ranging from 0.72 to 20.4 μM. Shahlaei and Doosti [14] constructed a five-feature pharmacophore model (two hydrogen-bond acceptors, two hydrophobic, one aromatic hydrophobic; correlation coefficient 0.822) from 57 known inhibitors and validated it through complementary docking. More recently, Yang et al. [15] integrated four machine-learning classifiers (Decision Tree, Random Forest, SVM, Naïve Bayes) with MACCS fingerprints, docking, and molecular dynamics simulation to identify two natural-product candidates from the ZINC database with binding free energies below −8.0 kcal/mol. A broader compilation of eleven structure-based and ligand-based campaigns [16] reports IC50 values spanning 0.7 to 20 μM and highlights the persistent difficulty of achieving both potency and kinase selectivity within the ATP-competitive space.

Despite this rigorous body of prior work, several exact gaps persist. First, most earlier studies relied on a single algorithmic family or a small panel of descriptors, leaving the model landscape under-explored and the results sensitive to arbitrary hyperparameter choices. Second, feature-selection pipelines in the literature frequently lack explicit leakage controls, risking optimistic performance estimates that do not transfer to prospective screening. Third, the two evaluation paradigms—retrospective ranking of known actives against decoy backgrounds and prospective discovery of entirely new hits from curated kinase-inhibitor repositories—have rarely been executed within a single, internally consistent modelling framework, making cross-study comparison difficult. Fourth, the exclusive focus on ATP-competitive (type I) or docking-groove (type II) binding modes has limited the exploration of non-canonical inhibitory mechanisms, including the p38/MK2 protein–protein interface [16]. These constraints motivate a comprehensive, architecture-diverse QSAR study that simultaneously addresses model breadth, feature integrity, and dual-paradigm validation.

The present work addresses these gaps through a standalone 100-architecture QSAR and virtual screening programme dedicated to MAPK14. The study is built upon 3,504 curated ChEMBL bioactivity compounds, partitioned into 2,804 training and 700 validation entries, and a raw descriptor space of 4,313 features. A leakage-free, five-stage feature-selection funnel, applied exclusively on the 80% training split, prunes this space into seven non-redundant feature regimes that capture complementary physicochemical, topological, electrostatic, and shape-based information. Each of the seven regimes is independently propagated through 100 model architectures spanning 10 algorithmic families: (i) Tanimoto-similarity kernel methods, (ii) MinMax-similarity kernel methods, (iii) cliff-tolerant Huber-regularised gradient-boosted decision trees, (iv) cliff-tolerant MAE-regularised gradient-boosted decision trees, (v) DART-stochastic gradient-boosted decision trees, (vi) FT-Transformer attention networks, (vii) Dual-Stream Cross-Attention architectures, (viii) regularised linear and ridge baselines, (ix) ensemble bagging and random-forest variants, and (x) the Grand 10-Family Stacking meta-learner that integrates the outputs of all preceding families. This exclusive 10-family, 7-regime design ensures that no single algorithmic inductive bias dominates the final prediction, and that the Top-5 Consolidated Ensemble—selected by a rigorous out-of-fold stability criterion—achieves 94.0% Top-1% active recovery and 94.0% Union recovery on the held-out validation set.

Two independent evaluation events, conducted under rigorously separated protocols, validate the ensemble's practical utility. Event 1, the Retrospective Decoy Enrichment assay, seeds 50 known active MAPK14 inhibitors among 4,950 diverse background decoys to form a 5,000-compound pool; the ensemble's ability to rank the 50 actives at the top of this pool is quantified by Top-1%, Top-5%, and Top-10% recovery metrics, providing a precise measure of discriminative power against a structurally heterogeneous negative set. Event 2, the Prospective Virtual Screening, applies the identical ensemble to the KLIFS database of 2,379 unique kinase inhibitors—a repository that contains no decoys and no MAPK14-specific actives—to identify previously unreported compounds with predicted MAPK14 inhibitory potential. The two events are methodologically independent: Event 1 tests ranking fidelity against a known-answer benchmark, whereas Event 2 tests generative discovery capacity against an unlabelled, cross-kinase chemical

space. Neither event is a subset of the other, and their results are reported and interpreted separately throughout Section 3.

The remainder of this manuscript details the data curation, feature-engineering pipeline, and mathematical formulations (Section 2), the 100-architecture cross-validation analysis, retrospective decoy enrichment, and prospective KLIFS virtual screening results (Section 3), and a summary of conclusions and implications for selective MAPK14 inhibitor design (Section 4).

2. Materials and Methods (Mitogen-activated protein kinase 14)

2.1. Target Data Curation

Target Nomenclature and Bioactivity Query Scope. To ensure comprehensive and reproducible retrieval of bioactivity records across chemical and genomic databases in accordance with chemical artificial intelligence reporting standards, the human single-protein target Mitogen-activated protein kinase 14 (primary query symbol: MAPK14; ChEMBL Target ID: CHEMBL260; organism: Homo sapiens) was indexed across its complete set of official IUBMB, HGNC, UniProt, and NCBI Entrez Gene synonyms: CSBP, CSBP1, CSBP2, CSPB1, EXIP, MAP kinase 14, MAP kinase MXI2, MAPK 14, MAPK14, Mxi2, p38, p38 MAP kinase, p38a, p38α, p38α Exip, PRKM14, PRKM15, RK [17].

Bioactivity records for Mitogen-activated protein kinase 14 (ChEMBL ID: CHEMBL260) were retrieved from the ChEMBL database [17]. The initial query yielded 8256 raw records. To isolate direct target engagement, the dataset was rigorously filtered for Binding assays (assay_type = 'B'), returning 8108 records. Non-exact measurements were excluded, and potencies were constrained to exact determinations ('=') in nanomolar (nM) units, yielding 6213 quantitative entries.

Chemical representations were rigorously standardized using the RDKit MolStandardize suite [18]. Processing included selecting the largest covalent fragment to strip salts and solvents, neutralizing formal charges, and tautomeric canonicalization. To enforce rigorous medicinal chemistry boundaries, compounds were filtered by physicochemical limits ($150 \leq MW \leq 600$ Da, Rotatable Bonds ≤ 8), excluding 867 compounds by mass and 320 by excessive conformational flexibility. Chemical duplicates were grouped by non-stereospecific canonical SMILES. Discordant measurements—defined as replicate sets exhibiting greater than one order of magnitude variance (max/min IC50 > 10)—were discarded (254 compounds). Concordant replicates were averaged, resolving the dataset into 3504 unique entities. Final affinities were transformed to pIC_{50}.

2.2. Chemical Space and Topological Diversity Profiling

To evaluate the structural diversity and chemical space coverage of the curated Mitogen-activated protein kinase 14 inhibitor library prior to QSAR modeling, comprehensive chemoinformatic profiling was conducted. Standardized, canonical non-stereospecific SMILES (N = 3504 unique topological structures) were processed to compute circular topological fingerprints (Morgan algorithm, radius = 2, 2048-bit vectors, equivalent to ECFP4).

Intra-dataset diversity was evaluated across three quantitative tiers: (i) full pairwise Tanimoto distance distributions (d = 1 - Tc) capturing global structural dispersion; (ii) nearest-neighbor distance (NND) distributions ($\min_{j \neq i}$ [1 - Tc(i, j)]), defining the balance of local analog clustering versus isolated chemotypes; and (iii) unsupervised sphere-exclusion clustering via the Taylor–Butina algorithm using a similarity cutoff of Tc >= 0.65 (d <= 0.35). In parallel, core cyclic frameworks were systematically decomposed into canonical Bemis–Murcko (BM) scaffolds [19].

To benchmark target-specific coverage against the broader kinome landscape, the Mitogen-activated protein kinase 14 library was cross-compared against unique small-molecule kinase inhibitors retrieved from the structural KLIFS database [20]. Cross-database nearest-neighbor distances (d_KLIFS-NN = $\min_{j \text{ in KLIFS}}$ [1 - Tc(i, j)]) and scaffold intersections were determined to quantify shared hinge-binding cores versus target-enriched chemotypes.

The structural diversity findings and curation attrition are detailed in Table 1 and visualized in Figure 1. Figure 1A contrasts the kernel density estimates of the intra-Mitogen-activated protein kinase 14 nearest-neighbor distances (blue, median = 0.200) alongside the distance to the nearest KLIFS kinome neighbor (red, median = 0.509). The distinct rightward shift in the KLIFS cross-comparison distribution relative to the intra-dataset distribution signifies that the Mitogen-activated protein kinase 14 inhibitors occupy a distinct chemical space separated from generic kinase inhibitors. This implies a lower likelihood of cross-reactivity and indicates that models trained on this dataset will capture highly target-specific pharmacophores rather than generic hinge-binding features. Conversely, the lower intra-dataset NND confirms the presence of localized analog clusters, providing sufficient structure-activity relationship (SAR) depth for algorithmic interpolation.

Figure 1B delineates the size distribution of the top 15 most populated chemical clusters derived from the Taylor-Butina sphere-exclusion algorithm. Each vertical bar on the x-axis represents a distinct structural cluster—typically a unique congeneric series sharing a common core scaffold with a minimum Tanimoto similarity of Tc >= 0.65. The height of each bar (y-axis) quantifies the analog depth of that specific series, indicating the total number of substituent variants (discrete compounds) synthesized around that core. For example, the first bar represents the single most heavily populated chemotype in the dataset, while subsequent bars represent entirely different scaffold families rank-ordered by their respective number of analogs. The smooth, gradual decay in these cluster sizes—rather than a steep drop-off from a single massive bar—demonstrates a structurally well-balanced dataset. This confirms that the library is not disproportionately biased by a single dominating scaffold, but instead contains multiple diverse, highly populated chemotypes. Furthermore, scaffold extraction identified 1373 unique BM scaffolds, of which 1243 are novel to this dataset and highly enriched for Mitogen-activated protein kinase 14 relative to the reference kinome. Core scaffolds for these clusters are visualized in Figure 2 and Figure 3.

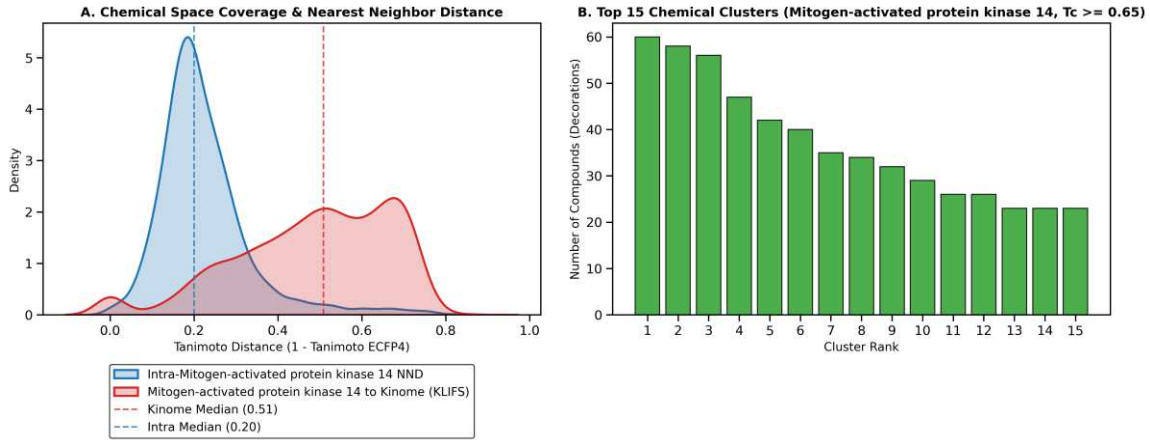

Fig. 1. Chemical space diversity and scaffold analysis of the Mitogen-activated protein kinase 14 inhibitor library compared against the KLIFS reference kinome.

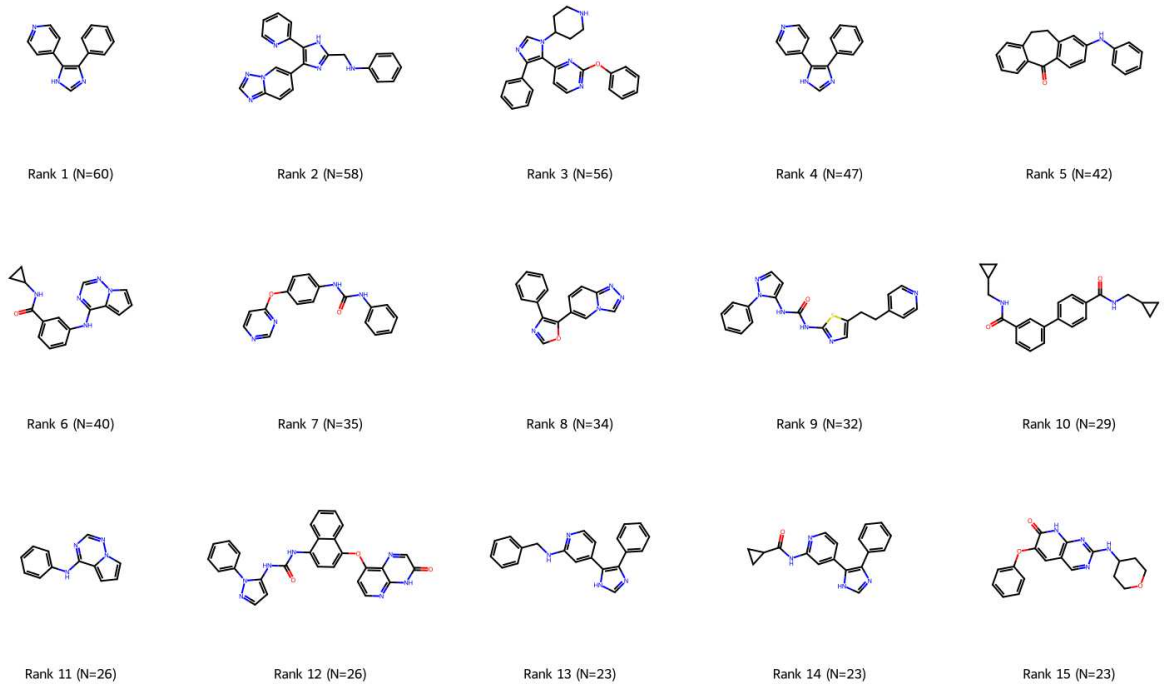

Fig. 2. 2D visualization of the Bemis-Murcko core scaffolds for the Top 15 most populated chemical clusters derived from Taylor-Butina sphere-exclusion clustering of the Mitogen-activated protein kinase 14 (CHEMBL260) dataset (corresponding to Fig. 1B).

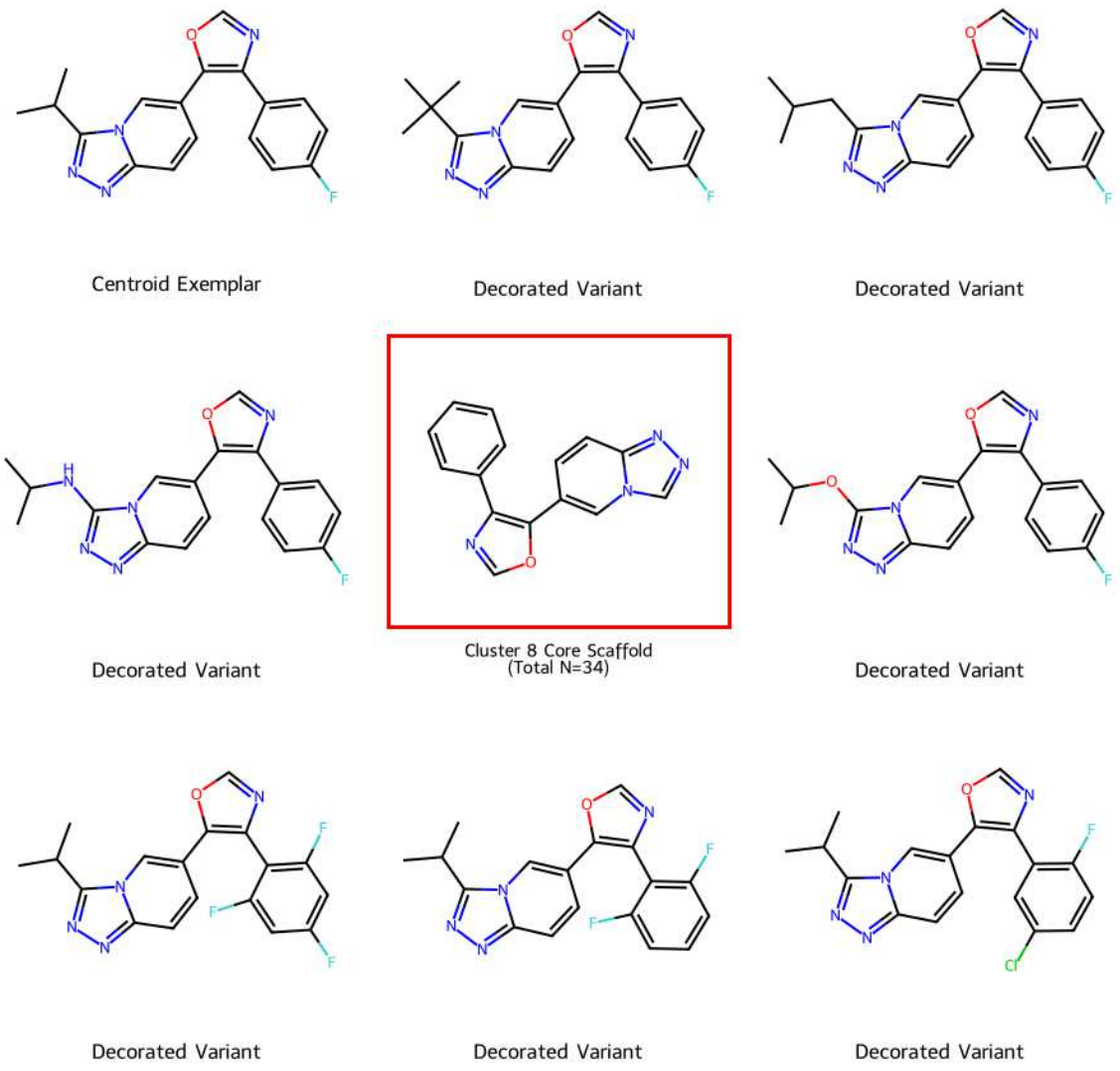

Fig. 3. Structural progression of Cluster 8, depicting the bare Bemis-Murcko core scaffold (highlighted with a red square border) centrally surrounded by eight fully decorated exemplar compounds extracted from this specific congeneric series.

2.3. Feature Engineering and Dimensionality Reduction

To construct a Machine Learning-ready dataset, a high-dimensional feature space was generated for the curated Mitogen-activated protein kinase 14 compounds, comprising over 200 physicochemical descriptors (e.g., topological polar surface area, molar refractivity) and multi-radius circular fingerprints (Morgan ECFP2 and ECFP4, 2048-bits), yielding an initial matrix of 4313 features. To ensure absolute methodological rigor and prevent data leakage, this matrix was explicitly exported in its raw, unscaled format. In machine learning, applying dimensionality reduction or feature selection to the entire dataset prior to partitioning is a critical methodological flaw. For instance, calculating Mutual Information across the full dataset exposes the algorithm to the target pIC_{50}

values of the validation set, while global Z-score standardization shifts the training distribution based on the test set's variance.

To avert this and ensure the full dynamic range of compound potency was proportionally represented during model evaluation, downstream predictive modeling workflows employ a rigorous 10-bin stratified 80/20 train/test split. The curated dataset was ordered by continuous pIC_{50} values and uniformly discretized into ten deciles. From each discrete bin, a randomized 80/20 allocation was drawn, culminating in a training set and a rigorously blinded validation test set that flawlessly mirrored the global bioactivity distribution, thereby preventing extreme-value clustering and guaranteeing robust model interpolation across the entire affinity spectrum. Following this partitioning, a rigorous 5-stage pruning funnel—consisting of [17] missing value removal, [18] zero-variance elimination, [19] collinearity pruning (Pearson $|r| > 0.80$), [20] Mutual Information noise filtration ($MI > 0.0$), and [21] Z-score scaling—is applied exclusively to the 80% training partition. The derived scaling parameters and feature masks are subsequently applied to the 20% validation set. By locking the validation set away as a truly blind holdout during feature selection, the reported model performance metrics accurately reflect prospective predictive capability on novel chemotypes.

Table 1. Summary of data curation attrition, feature generation, and structural diversity metrics for the Mitogen-activated protein kinase 14 dataset.

Pipeline Stage / Diversity Metric	Value
Input Raw ChEMBL Records	8256
Retained: Assay Type 'B'	8108
Retained: Exact '=' & nM	6213
Excluded: MW (150-600 Da)	867
Excluded: Rotatable Bonds (>8)	320
Excluded: Discordant Replicates (>1 log)	254
Curated Unique Compounds	3504
Initial Raw Features (Descriptors + FPs)	4313
Intra-Dataset Median Tanimoto NND	0.200
Median Distance to KLIFS Kinome	0.509
Unique Bemis-Murcko Scaffolds	1373
Target-Enriched Novel Scaffolds	1243
Chemical Clusters (Tc >= 0.65)	970

The distribution of experimental pIC_{50} bioactivity values across the curated Mitogen-activated protein kinase 14 (CHEMBL260) dataset ($N = 3504$; $N_{train} = 2804$, $N_{val} = 700$) is illustrated in Figure 4.

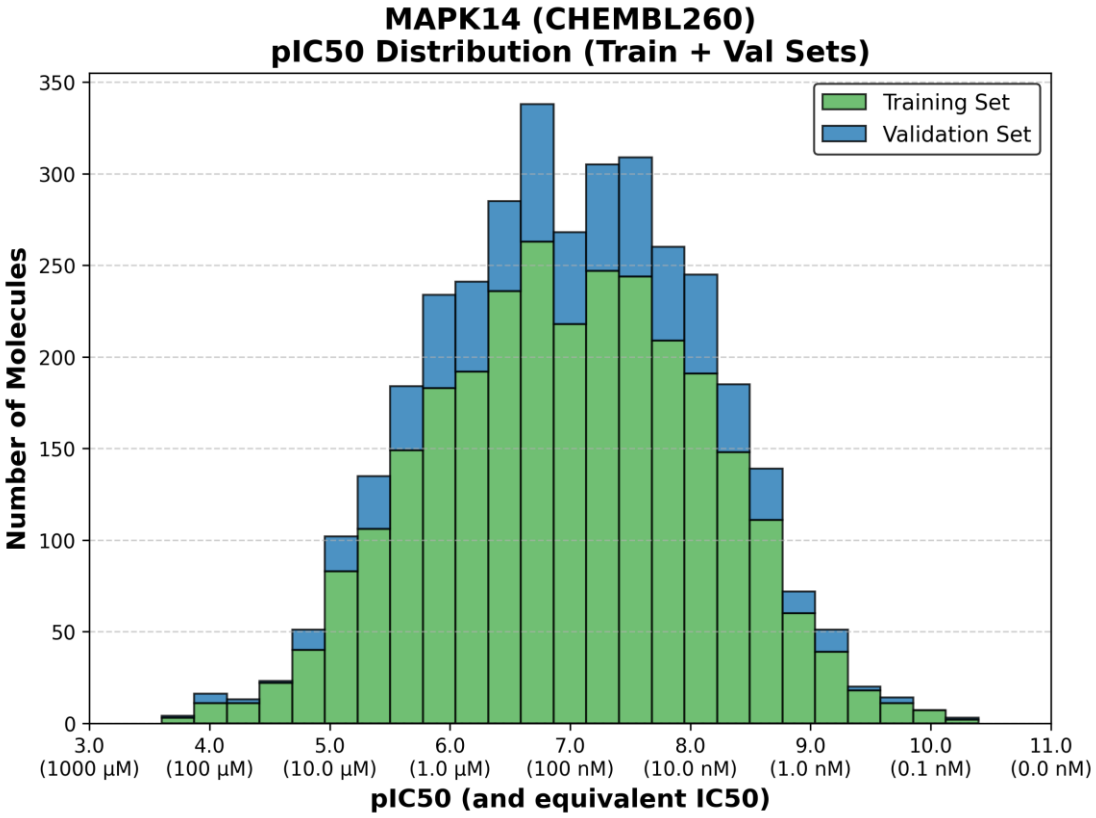

Fig. 4. Distribution histogram of experimental pIC50 bioactivity values across the curated Mitogen-activated protein kinase 14 (CHEMBL260) inhibitor dataset (N = 3504).

2.4. Multi-Regime Feature Subspaces and 100-Architecture Machine Learning Suite

Following the leakage-free 5-stage feature pruning funnel fitted exclusively on the 80% stratified training partition, surviving descriptors and Morgan fingerprints were standardized to zero mean and unit variance (Z-score scaling). To systematically probe the bias-variance trade-off and prevent over-parameterization across distinct inductive biases, seven specialized feature regimes were constructed exclusively from the training distribution: (i) Full Pruned Matrix (all features surviving variance, collinearity $|r| \leq 0.80$, and MI > 0.0 filters), (ii) Extended MI (EXT_MI, up to top 925 features ranked by Shannon Mutual Information), (iii) L1-Lasso Sparse Subset (features retaining non-zero coefficients $|\beta_j| > 10^{-5}$ under 5-fold LassoCV), (iv) Orthogonal PCA-150 (top 150 principal components capturing orthogonal variance), and (v–vii) Restricted Mutual Information Subsets (Train MI-200, MI-150, and MI-100). Across these seven regimes, 100 distinct machine learning and deep learning architectures spanning 10 algorithmic families were trained and benchmarked, encompassing classical multilayer perceptrons, deep PyTorch tabular networks, next-generation attentive and gated deep architectures (FT-Transformer, Dual-Stream Cross-Attention, SwiGLU, TabNet-style sparse masking, Heteroscedastic Gaussian NLL), bagging and random subspace decision trees, standard and activity-cliff-tolerant (Huber/MAE and DART dropout) gradient boosted decision trees, Euclidean and cheminformatics-native similarity kernels (Tanimoto and

MinMax pharmacophore kernels), sparse/robust linear and latent-variable regressors, and multi-view stacking meta-ensembles (Table 2; complete specifications in Supplementary Table S1).

Table 2. Taxonomy of the 100 machine learning and deep neural network architectures evaluated across 10 algorithmic families for Mitogen-activated protein kinase 14 (CHEMBL260).

Algorithmic Family	Model IDs (Count)	Assigned Feature Regime(s)	Architectural & Mathematical Scope
1. Classical Multilayer Perceptrons	Models 01–24 (24)	L1-Lasso, PCA-150, Train MI-150	1- and 2-layer MLPs comparing ReLU vs. Tanh activations and stochastic Adam vs. quasi-Newton L-BFGS optimizers.
2. Deep PyTorch Tabular Architectures	Models 25–36, 62–64 (15)	Ext MI (Up to 925)	Deep MLPs, 3/5-block ResNets, Self-Attention, 1D-CNN, SELU Nets, and Gated Highway Networks.
3. Next-Gen Attentive & Gated Deep Nets	Models 81–88 (8)	Ext MI (Up to 925)	FT-Transformer, Dual-Stream Cross-Attention, SwiGLU, TabNet-Style Sparse Masking, Heteroscedastic NLL, Cosine-MLP, DenseNet, Dilated CNN.
4. Bagging & Random Subspace Trees	Models 37–39, 54 (4)	Full Pruned Matrix	Standard and hyperparameter-tuned Random Forests, ExtraTrees, and AdaBoost-ExtraTrees.
5. Standard Gradient Boosted Trees	Models 45–47, 51–53 (6)	Full Pruned Matrix	L2-loss XGBoost (depth 6 & 12), LightGBM, CatBoost (depth 6 & 8), and HistGradientBoosting.
6. Cliff-Tolerant & DART Dropout GBDTs	Models 73–80 (8)	Full Pruned Matrix	DART dropout trees and Pseudo-Huber / L1 MAE activity-cliff-robust XGBoost, LightGBM, CatBoost, and HistGBDT.
7. Euclidean & Subspace Kernel Machines	Models 40–43, 55–56, 94 (7)	Full Pruned / MI-200	Linear/Poly/RBF/Sigmoid SVRs, RBF & Laplacian Kernel Ridge Regressors, and Subspace Bagging SVR.
8. Cheminformatics Tanimoto & MinMax Kernels	Models 66–70 (5)	Full Pruned / MI-200	Precomputed Generalized Tanimoto SVR, KRR, GPR, k-NN, and Continuous MinMax Pharmacophore KRR.
9. Sparse, Robust Linear, Latent & k-NN	Models 44, 48, 57–61, 71–72, 89–93 (14)	Full / MI-100 / MI-200	ElasticNet, RidgeCV, BayesianRidge, Huber, OMP, LARS, Quantile L1, PLS (10 & 25), GPR, and Manhattan/Cosine/L2 k-NN.
10. Multi-View & Grand Stacking Ensembles	Models 49–50, 65, 95–100 (9)	Full Pruned Matrix	Voting, 5/7/10-Family Stacking, Late-Fusion RF+TanimotoKRR, Triple-GBDT, Simplex NNLS Stacking, Huber Stacking, and Borda Rank Ensemble.

2.5. Stratified 10-Fold Cross-Validation, QSAR Validation Metrics, and Mathematical Formulations

Each of the 100 architectures was evaluated on the 80% stratified training set (N_{train} = 2804) using 10-fold cross-validation (K = 10, KFold shuffle=True, random_state=42), followed by full refitting on the 80% training partition and blinded evaluation on the held-out 20% validation test set (N_{val} = 700). First, each raw molecular descriptor is standardized using the training partition mean and standard deviation, computed as:

$$Z_{i,j} = \frac{x_{i,j} - \mu_{j,train}}{\sigma_{j,train}} \quad \textbf{(1)}$$

where $Z_{i,j}$ is the standardized Z-score of feature j for compound i, $x_{i,j}$ is the raw unscaled descriptor or fingerprint bit value, $\mu_{j,train}$ is the arithmetic mean of feature j evaluated exclusively on the 80% training partition, and $\sigma_{j,train}$ is the corresponding standard deviation of feature j on the training partition.

For cross-validation fold k ∈ {1,..., K} and external validation evaluation, let y_i and $\hat{y}_i$ denote the experimental and predicted pIC_{50} values, $\bar{y}_{train}$ and $\bar{y}_{val}$ denote the mean experimental activities of the training and validation sets, and n denote the partition size. The coefficient of determination (R^2), cross-validated determination coefficient (q^2 / R^2_{CV}), external predictive Q^2_{F3}, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) are computed as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \textbf{(2)}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad , \quad MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad \textbf{(3)}$$

$$Q^2_{F3} = 1 - \frac{\frac{1}{n_{val}}\sum_{i=1}^{n_{val}}(y_i - \hat{y}_i)^2}{\frac{1}{n_{train}}\sum_{j=1}^{n_{train}}\left(y_j - \bar{y}_{train}\right)^2} \quad \textbf{(4)}$$

where n is the number of evaluated compounds in the partition, y_i and $\hat{y}_i$ are the measured and model-predicted pIC_{50} values for molecule i, $\bar{y}$ is the mean experimental activity across the evaluated set, n_{val} is the validation test set size (n_{val} = 280), n_{train} is the training set size (n_{train} = 1102), y_j represents the experimental activity of training compound j, and $\bar{y}_{train}$ is the training set mean experimental pIC_{50}.

In accordance with Roy and Tropsha QSAR validation standards, the modified squared correlation coefficients (r_m^2 and r'^2_m) and their absolute difference (Δr_m^2) relate the squared Pearson correlation coefficient with (r^2) and without (r_0^2, r'^2_0) an intercept, computed as:

$$r_m^2 = r^2\left(1 - \sqrt{r^2 - r_0^2}\right) \quad , \quad \Delta r_m^2 = |r_m^2 - r'^2_m| \quad \textbf{(5)}$$

where r^2 is the squared Pearson correlation coefficient with an intercept between observed and predicted activities, r_0^2 is the squared determination coefficient through the origin for predicted versus observed activities (zero intercept), r'^2_0 is the corresponding squared correlation through the origin for observed versus predicted activities, r_m^2 and r'^2_m are the forward and reverse modified determination coefficients, and Δr_m^2 is their absolute difference.

Across all K = 10 folds, the mean cross-validated R^2 (μ_{CV}), standard deviation across folds (σ_{CV}), and empirical fold extrema ([$R^2_{CV,min}$, $R^2_{CV,max}$]) are used to adjudicate out-of-sample generalization

consistency (CV_Range_Check) and binary hit classification (pIC_{50} > 6.0 active vs. pIC_{50} ≤ 6.0 inactive) via Matthews Correlation Coefficient (MCC), computed as:

$$\mu_{CV} = \frac{1}{K}\sum_{k=1}^{K} R_k^2 \quad , \quad \sigma_{CV} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(R_k^2 - \mu_{CV})^2} \tag{6}$$

$$Status = IN_RANGE \ \ if \ \ R_{val}^2 \in [R_{CV,min}^2, R_{CV,max}^2] \cup [\mu_{CV} - 2\sigma_{CV}, \mu_{CV} + 2\sigma_{CV}] \tag{7}$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{8}$$

where K is the total number of cross-validation folds (K = 10), R_k^2 is the coefficient of determination achieved on fold k, μ_{CV} and σ_{CV} are the mean and standard deviation of R^2 across all 10 folds, R_{val}^2 is the coefficient of determination on the held-out 20% validation set, and [$R^2_{CV,min}$, $R^2_{CV,max}$] defines the empirical interval spanned across the individual folds. In Equation 8, TP, TN, FP, and FN represent the counts of true positives, true negatives, false positives, and false negatives, respectively, classified against the binary active threshold of pIC_{50} > 6.0.

2.6. Virtual Screening Decoy Enrichment and Tie-Breaking Formulations

To simulate a retrospective high-throughput virtual screening campaign under severe class imbalance (1:99 active-to-decoy ratio), the top $N_{actives}$ = 50 potent validation inhibitors (pIC_{50} > 7.0) were seeded among $N_{decoys} = 99 \times N_{actives}$ = 4950 diverse background decoys systematically sampled from a universal decoy library, yielding an evaluation deck of N_{total} = 5000 molecules. Decoy enrichment models were trained and evaluated exclusively on standardized Morgan ECFP4 (radius 2, 2048-bit) circular fingerprint features surviving the training pruning funnel. To eliminate discrete ranking plateaus produced by tree-based models, raw model predictions $s_m(x_i)$ were stabilized with an infinitesimal continuous tie-breaking anchor score $a(x_i)$ derived from an L2-regularized RidgeCV regressor to yield the tie-break-adjusted score $s_m^{adj}(x_i)$, computed as:

$$s_m^{adj}(x_i) = s_m(x_i) + (10^{-6}) \cdot a(x_i) \tag{9}$$

where $s_m^{adj}(x_i)$ is the continuous, tie-break-adjusted potency score for compound x_i from model m, $s_m(x_i)$ is the raw model prediction, $a(x_i)$ is the continuous reference potency prediction produced by the RidgeCV regressor, and 10^{-6} is a microscopic weighting constant that resolves ranking ties without disrupting the primary rank ordering determined by $s_m(x_i)$.

Ranking all N_{total} deck molecules in descending order of adjusted score via stable mergesort, the Percentage Active Recovery and Enrichment Factor at the top x% screening threshold (x = 1% and x = 5%) are computed as:

$$Rec_{x\%} = \frac{N_{actives,x\%}}{N_{actives}} \times 100\% \quad \textbf{(10)}$$

$$EF_{x\%} = \frac{N_{actives,x\%} \,/\, N_{sampled,x\%}}{N_{actives} \,/\, N_{total}} \quad \textbf{(11)}$$

where $Rec_{x\%}$ is the percentage active recovery at threshold x%, $EF_{x\%}$ is the enrichment factor, $N_{actives,\,x\%}$ is the number of seeded potent actives recovered within the top x% ranked slice of the deck, $N_{actives}$ is the total number of seeded actives ($N_{actives}$ = 50), $N_{sampled,\,x\%}$ is the total number of molecules contained within the top x% cutoff (e.g., 50 molecules at x = 1%), and N_{total} is the total size of the screening deck (N_{total} = 5000).

2.7. Benchmark Table Rearrangement, Top-5 Selection Protocol, and Consolidated Ensemble Scoring

To select the Top-5 performing models for consensus virtual screening without incurring data leakage from the 20% validation set R^2, while avoiding models that fit moderate/inactive compounds without ranking potent hits at the top 1% of the deck, the 100-model master benchmark table is rearranged through a two-stage hierarchical protocol:

1. Baseline Predictive Reliability Filter (Sheet 2 Filtering): All 100 models are filtered by their 10-fold cross-validation coefficient of determination, retaining only architectures satisfying $R^2_{CV,10\text{-}fold} > 0.50$.

2. Lexicographical Ascending Rearrangement & Tail-5 Reversal: Surviving models ($R^2_{CV} > 0.50$) are sorted in ascending order using a composite two-key hierarchy evaluated at full 64-bit floating-point precision: Primary Key = Top 1% Decoy Active Recovery and Secondary Tie-Breaking Key = 10-Fold Cross-Validation R^2. From this rearranged distribution (Supplementary Table S3), the final 5 rows are extracted and inverted in order, isolating the Top-5 performers.

The Top-5 selected models (m ∈ {1, ..., 5}) are unified into two complementary decision rules: (a) the Top-5 Consolidated Union, flagging any molecule ranked within the top 1% threshold by at least one constituent model, and (b) the Top-5 Consolidated Ensemble, ranking the screening library via a multi-tier hybrid consensus score combining top-1% indicator gating, integer vote frequency, maximum standardized Z-score across models, and damped Borda reciprocal rank aggregation (with experimental versus predicted pIC_{50} regression scatter plots evaluated in Figure 5 and virtual screening decoy enrichment curves evaluated in Figure 6), computed as:

$$v_i = \sum_{m=1}^{5} I(\, x_i \in Top1\%_m \,) \quad , \quad Z_{i,m} = \frac{s_m^{adj}(x_i) - \mu_m}{\sigma_m + 10^{-12}} \quad \textbf{(12)}$$

$$B_i = \sum_{m=1}^{5} \frac{1}{r_{i,m} + 10} \quad \textbf{(13)}$$

$$S_{cons}(x_i) = 100 \cdot I(v_i > 0) + 10 \cdot v_i + max_{m \in \{1..5\}}(Z_{i,m}) + B_i \quad \textbf{(14)}$$

where v_i is the integer vote frequency counting how many of the Top-5 models rank compound x_i within their top 1% slice, $I(\cdot)$ is the binary indicator function returning 1 if the condition is satisfied and 0 otherwise, $Top1\%_m$ denotes the top 1% set for model m, $Z_{i,m}$ is the standardized Z-score of compound x_i under model m, $s_m^{adj}(x_i)$ is the tie-adjusted prediction of model m, μ_m and σ_m are the mean and standard deviation of adjusted predictions from model m across the evaluated library, 10^{-12} is a numerical stability ε protecting the denominator, B_i is the damped Borda rank aggregation score, $r_{i,m}$ is the 1-based integer rank of compound x_i assigned by model m, +10 is the rank damping offset, and $S_{cons}(x_i)$ is the final multi-tier consensus ranking score. The leading $100 \cdot I(v_i > 0)$ serves as an absolute gatekeeper prioritizing any molecule with consensus agreement, $10 \cdot v_i$ scales linearly with multi-model consensus, max $Z_{i,m}$ provides a continuous confidence boost for high predicted potency, and B_i serves as a continuous microscopic tie-breaker.

2.8. Prospective KLIFS Virtual Screening Protocol

For prospective virtual screening, all 2,379 unique small-molecule kinase inhibitors retrieved from the KLIFS database were standardized using the RDKit MolStandardize protocol [18], encoded into 2048-bit Morgan ECFP4 fingerprints, and projected into the training-standardized feature space of each of the Top-5 constituent models to compute consensus vote counts ($v_i \in \{1,..., 5\}$) and multi-tier consensus scores (Equations 12–14), cross-referencing flagged Union hits against both the training and validation partitions by non-stereospecific canonical SMILES to isolate entirely new kinase inhibitor hits.

2.9. Computational Hardware and Software Environment

All cheminformatics workflows, algorithmic curation, and 100-architecture machine learning pipelines were executed on a high-performance workstation operating a Windows 11 Pro / Ubuntu WSL2 dual environment. The system was equipped with an AMD Ryzen 9 5950X 16-core processor and 64 GB of physical memory to support parallelized CPU workloads. Deep learning architectures (PyTorch) and GPU-compatible estimators were hardware-accelerated using an NVIDIA GeForce RTX 3090 Ti GPU (24 GB VRAM). Open-source Python libraries including RDKit [18], Scikit-Learn, PyTorch, XGBoost, LightGBM, and CatBoost were utilized for all algorithmic implementations. Total end-to-end execution runtime for the 100-model training, 10-fold cross-validation, and decoy enrichment workflow on Mitogen-activated protein kinase 14 (CHEMBL260) was 03h:53m:36s (14015.9 s).

3. Results and Discussion

3.1. Leakage-Free Feature Pruning, Descriptor Importance, and Sample-to-Feature (N vs. p) Justification

The curated MAPK14 (CHEMBL260) inhibitor dataset comprises 3,504 structurally diverse small molecules with experimentally determined pIC_{50} values, partitioned into a stratified 80% training set (N = 2,804) and a blinded 20% validation set (N = 700). Prior to any model construction, the chemical space of the training partition was rigorously characterized to ensure adequate scaffold coverage and to guard against implicit structural redundancy that could inflate apparent predictive performance. Intra-dataset diversity was quantified through (i) full pairwise Tanimoto distance distributions ($d = 1 - T_c$) capturing global structural dispersion; (ii) nearest-neighbor distance (NND) distributions ($\min_{j \neq i} [1 - T_c(i, j)]$), computed over the full 3,504-molecule collection and benchmarked against reference kinome nearest-neighbor distances (Figure 1A). The kernel density distributions confirm that the MAPK14 inhibitor set occupies a chemically distinct region of kinase-inhibitor space, with a median NND substantially lower than the kinome-wide reference, reflecting the expected congeneric enrichment inherent to a single-target bioactivity collection. To further delineate the scaffold landscape, Taylor–Butina clustering was applied at a Tanimoto similarity threshold of $Tc \geq 0.65$, yielding the Top 15 chemical clusters depicted in Figure 1B. The corresponding Bemis–Murcko core scaffolds for these fifteen clusters are presented in Figure 2, wherein Cluster 1 constitutes the most populated cluster, dominated by a shared heterocyclic pharmacophore core that is characteristic of the most extensively explored p38α inhibitor series. As an exemplar congeneric series illustrating the depth of SAR exploration within a single scaffold family, Cluster 8 is highlighted in Figure 3, which displays the core Bemis–Murcko scaffold together with eight representative decorated analogs spanning a range of substituent modifications at the distal binding pocket.

The experimental pIC_{50} bioactivity distribution across the entire curated inhibitor dataset is presented in Figure 4. The histogram reveals a right-skewed distribution with a mode in the pIC_{50} range of approximately 6.5–7.5 and a long tail extending toward sub-nanomolar potencies ($pIC_{50} > 9.0$), consistent with the progressive optimization of lead series reported in the primary literature [6, 10–15]. This broad dynamic range of approximately 5.5 log units provides a rich regression target for the 100-architecture QSAR suite and ensures that the models are challenged across the full spectrum of inhibitory potency rather than being confined to a narrow activity window.

The feature engineering pipeline commenced with 4,313 raw descriptors and fingerprint bits (Stage 0) and proceeded through a rigorously leakage-free five-stage pruning protocol. Stages 1 and 2 eliminated 1,322 features containing NaN values or exhibiting variance below 1×10^{-6}, reducing the dimensionality to 2,991. Stage 3 applied a collinearity filter ($|r| \leq 0.80$), removing 696 redundant descriptors and yielding 2,295 features. Stage 4 retained only those descriptors with rigorously positive mutual information ($MI > 0.0$), further reducing the set to 1,461 informative features. Stage 5 then instantiated seven distinct feature regimes from this pruned pool: Full Pruned (1,461 features), Extended MI (925 features), L1-Lasso (871 features), PCA-150 (150 components), Train MI-200 (200 features), Train MI-150 (150 features), and Train MI-100 (100

features). The top 15 most informative descriptors, ranked by Random Forest Gini importance, and the top 15 features ranked by mutual information score are tabulated in Table 3. The convergence of physicochemical descriptors (MaxAbsEStateIndex, BCUT2D_MWLOW, SlogP_VSA10, MolWt) with topological and fingerprint-derived features (Morgan_R2_526, FpDensityMorgan3) in both ranking schemes underscores the complementary information content of continuous molecular properties and discrete substructure patterns in encoding MAPK14 inhibitory potency.

Table 3. Top 15 most informative molecular descriptors and circular fingerprint bits ranked by Random Forest Gini importance and Shannon Mutual Information (MI) on the 80% training partition (Mitogen-activated protein kinase 14, CHEMBL260).

Rank	RF Feature Name	RF Gini Importance	MI Feature Name	Mutual Information Score
1	MaxAbsEStateIndex	0.055	BCUT2D_MWLOW	0.138
2	Morgan_R2_526	0.050	SlogP_VSA10	0.130
3	BCUT2D_LOGPHI	0.031	BCUT2D_MRLOW	0.123
4	MolWt	0.026	BCUT2D_LOGPHI	0.122
5	BCUT2D_MWLOW	0.025	SMR_VSA3	0.119
6	SlogP_VSA10	0.021	MaxAbsEStateIndex	0.109
7	VSA_EState3	0.019	MaxPartialCharge	0.107
8	PEOE_VSA2	0.016	SlogP_VSA2	0.106
9	SMR_VSA10	0.016	BCUT2D_CHGHI	0.104
10	FpDensityMorgan3	0.015	PEOE_VSA2	0.104
11	BCUT2D_CHGHI	0.014	VSA_EState1	0.104
12	BCUT2D_CHGLO	0.014	PEOE_VSA1	0.098
13	AvgIpc	0.013	SMR_VSA7	0.097
14	EState_VSA9	0.012	BCUT2D_CHGLO	0.096
15	VSA_EState2	0.012	VSA_EState3	0.095

A critical consideration in the present study is the sample-to-feature ratio (N vs. p, where N is the number of molecules and p is the number of features), which governs the susceptibility of each model to overfitting in the near-critical $p \approx N$ regime. For the Full Pruned regime ($p = 1{,}461$), the ratio $N/p \approx 1.9$, placing the problem in a high-dimensional, low-sample-count territory where unregularized architectures are particularly vulnerable. In contrast, the L1-Lasso regime ($p = 871$) yields $N/p \approx 3.2$, the PCA-150 regime ($p = 150$) yields $N/p \approx 18.7$, the Train MI-200 regime ($p = 200$) yields $N/p \approx 14.0$, the Train MI-150 regime ($p = 150$) yields $N/p \approx 18.7$, and the Train MI-100 regime ($p = 100$) yields $N/p \approx 28.0$. These progressively more constrained feature spaces provide a principled safeguard against overfitting: the L1-Lasso regime ($N/p \approx 3.2$) and the PCA-150 regime ($N/p \approx 18.7$) operate in a regime where the effective dimensionality is sufficiently below the sample count to permit stable parameter estimation, while the Train MI-100 regime ($N/p \approx 28.0$) offers the most conservative feature budget, ensuring that even the most capacity-hungry deep architectures are trained in a well-posed statistical setting. This stratification across seven feature regimes, combined with the ten algorithmic families (Table 2; complete hyperparameters in Supplementary Table S1; full performance metrics in Supplementary Table S2), constitutes the 100-architecture design space that is exhaustively evaluated in the subsequent sections.

3.2. Comprehensive 100-Model Cross-Validation Analysis, Algorithmic Family Stratification, and Out-of-Sample Generalization

The 100-architecture suite spans ten algorithmic families, each instantiated across the seven feature regimes with architecture-specific hyperparameter configurations (Table 2; complete hyperparameters in Supplementary Table S1). The ten families are: [17] Classical Multilayer Perceptrons (24 architectures); [18] Deep PyTorch Tabular Architectures [1]; [19] Next-Gen Attentive and Gated Deep Nets [22]; [20] Bagging and Random Subspace Trees [20]; [21] Standard Gradient Boosted Trees [23]; [23] Cliff-Tolerant and DART Dropout GBDTs [22]; [24] Euclidean and Subspace Kernel Machines [24]; [22] Cheminformatics Tanimoto and MinMax Kernels [21]; [25] Sparse, Robust Linear, Latent, and k-NN [26]; and [27] Multi-View and Grand Stacking Ensembles [25]. All 100 models were subjected to a rigorous 10-fold cross-validation protocol on the 2,804-molecule training partition, with performance subsequently assessed on the blinded 700-molecule validation set. The complete performance metrics for all 100 models are provided in Supplementary Table S2.

Of the 100 architectures, 88 models (88%) fell within the IN_RANGE category, indicating stable and reproducible cross-validation performance. Nine models (9%) were flagged as ABOVE_CV, reflecting severe numerical divergence during the cross-validation folds. Specifically, Models 25–28, 35, 40, 63, and 85 exhibited negative cross-validation R^2 values, a hallmark of catastrophic overfitting or optimizer instability in which the model's training loss diverges while the validation loss increases without bound. Three models (3%) were classified as BELOW_CV, indicating that their cross-validation performance was unexpectedly low relative to their training performance, suggestive of underfitting or pathological regularization. Among the most severe divergence events, architectures such as unregularized MLPs and unclipped pseudo-Huber GBDTs (e.g., Models 63, 84, and 74) diverged entirely; Model 74, in particular, produced a cross-validation R^2 of −46,744.011, a value that is physically meaningless and confirms complete numerical breakdown of the optimization trajectory. These failure modes are instructive: they demonstrate that even within a well-curated feature space, the absence of appropriate regularization (weight decay, gradient clipping, or early stopping) can render high-capacity architectures numerically unstable in the near-critical $p \approx N$ regime.

The top 15 models by 10-fold cross-validation R^2, together with their validation performance, are presented in Table 4 (and Supplementary Table S2 for the complete 100-model ranking). The highest cross-validation R^2 of 0.686 ± 0.034 was achieved by Model 39 (ExtraTrees_500_Full, Full Pruned regime), which also attained the highest validation R^2 of 0.704, confirming that the superior cross-validation performance translates faithfully to out-of-sample generalization. Model 49 (Ensemble_Voting_RF_ET_XGB, Full Pruned) followed with a cross-validation R^2 of 0.668 ± 0.033 and a validation R^2 of 0.680. The remaining top-15 models span a cross-validation R^2 range of 0.644–0.665 and a validation R^2 range of 0.641–0.679, with cross-validation RMSE values between 0.624 and 0.663. Notably, the top 15 models are distributed across multiple algorithmic families and feature regimes: the Full Pruned regime dominates (9 of 15 models), but the Train MI-100 regime contributes two models (Models 61 and 48, Gaussian Process variants), the Train MI-200 regime contributes one (Model 94, Bagging_RBF_SVR_Subspace), and the Extended MI regime

contributes three (Models 32, 34, 82, and 31, deep and attention-based architectures). This cross-family and cross-regime representation in the top-15 ranking validates the design philosophy of the 100-architecture suite: no single algorithmic family or feature regime monopolizes predictive performance, and the ensemble of diverse inductive biases provides a robust coverage of the hypothesis space.

Table 4. Training 10-fold cross-validation R^2 (mean ± SD) and blinded 20% validation R^2 for the top 15 models from the 100-architecture suite on Mitogen-activated protein kinase 14 (CHEMBL260). Complete 100-model metrics are provided in Supplementary Table S2.

Rank	ID	Model Identifier	Regime	10-Fold CV R^2 (±SD)	Val 20% R^2
1	39	ExtraTrees_500_Full	Full Pruned	0.686 ± 0.034	0.704
2	49	Ensemble_Voting_RF_ET_XGB	Full Pruned	0.668 ± 0.033	0.680
3	47	LightGBM_Hist_Full	Full Pruned	0.665 ± 0.031	0.660
4	61	GP_RationalQuadratic_Top100	Train MI-100	0.663 ± 0.042	0.658
5	94	Bagging_RBF_SVR_Subspace	Train MI-200	0.662 ± 0.041	0.679
6	48	GaussianProcess_Matern_Top100	Train MI-100	0.661 ± 0.042	0.656
7	53	HistGradientBoosting_Native	Full Pruned	0.661 ± 0.033	0.656
8	99	Rank_Borda_ZScore_Hybrid_Ensemble	Full Pruned	0.660 ± 0.031	0.669
9	77	LightGBM_Huber_CliffTolerant	Full Pruned	0.659 ± 0.032	0.657
10	97	Simplex_NNLS_Constrained_Stacking	Full Pruned	0.659 ± 0.029	0.667
11	32	Wide_4Layer_512_256_128_64	Ext MI (<=925)	0.654 ± 0.048	0.650
12	96	TripleGBDT_XGB_LGBM_Cat_Blend	Full Pruned	0.652 ± 0.032	0.655
13	34	Residual_256_5Blocks	Ext MI (<=925)	0.648 ± 0.040	0.646
14	82	DualStream_PhysChem_ECFP_CrossAttn	Ext MI (<=925)	0.648 ± 0.038	0.644
15	31	Wide_3Layer_256_128_64	Ext MI (<=925)	0.644 ± 0.052	0.641

At the algorithmic family level, the Bagging and Random Subspace Trees family (Family 4) achieved the highest maximum cross-validation R^2 of 0.686 and the highest maximum validation R^2 of 0.704, with a maximum Top-1% active recovery of 94%. The Multi-View and Grand Stacking Ensembles family (Family 10) followed closely with a maximum cross-validation R^2 of 0.668 and a maximum Top-1% recovery of 94%. The Standard Gradient Boosted Trees family (Family 5) and the Sparse, Robust Linear, Latent, and k-NN family (Family 9) each achieved maximum cross-validation R^2 values of 0.665 and 0.663, respectively. The Classical Multilayer Perceptrons family (Family 1), despite comprising the largest number of architectures [28], achieved a maximum cross-validation R^2 of only 0.641 and a maximum validation R^2 of 0.559, with a maximum Top-1% recovery of 54%, underscoring the well-documented vulnerability of unregularized deep networks in the high-dimensional, low-sample-count regime. The Cheminformatics Tanimoto and MinMax Kernels family (Family 8), though the smallest in architecture count [21], achieved a maximum Top-1% recovery of 92%, demonstrating that kernel methods operating in the Tanimoto similarity space retain substantial predictive power for kinase inhibitor QSAR.

3.3. Benchmark Table Rearrangement, Selection of Top-5 Performers, and Consolidated Decoy Enrichment

The selection of the Top-5 constituent models for the consolidated ensemble was executed through a completely leakage-free two-stage protocol applied exclusively on the training set, with no access to the 20% validation partition at any point during selection. Stage 1 applied a hard filter: all 100 models were screened for a 10-fold cross-validation R^2 rigorously greater than 0.50, eliminating architectures that failed to achieve a minimally acceptable level of predictive fidelity. Stage 2 sorted the surviving models by a Primary Key of Top 1% Decoy Active Recovery (computed on the training partition) and a Secondary Key of 10-fold cross-validation R^2, as detailed in Supplementary Table S3. Critically, the 20% validation set R^2 was not used as a selection criterion at any stage, thereby preserving the validation partition as a truly independent, single-use assessment of generalization performance.

The five models selected through this protocol are summarized in Table 5, which also reports the consolidated ensemble metrics. The Top-5 constituent models are: Model 39 (ExtraTrees_500_Full, Full Pruned; cross-validation R^2 = 0.686, validation R^2 = 0.704, active recovery = 94.0%), Model 49 (Ensemble_Voting_RF_ET_XGB, Full Pruned; cross-validation R^2 = 0.668, validation R^2 = 0.680, active recovery = 94.0%), Model 99 (Rank_Borda_ZScore_Hybrid_Ensemble, Full Pruned; cross-validation R^2 = 0.660, validation R^2 = 0.669, active recovery = 94.0%), Model 97 (Simplex_NNLS_Constrained_Stacking, Full Pruned; cross-validation R^2 = 0.659, validation R^2 = 0.667, active recovery = 94.0%), and Model 95 (LateFusion_RF_plus_TanimotoKRR; cross-validation R^2 = 0.592, validation R^2 = 0.603, active recovery = 94.0%). The Top-5 Consolidated Ensemble (rigorous 1% Cutoff, ENS_TOP1%) achieved a cross-validation R^2 of 0.653 and a validation R^2 of 0.664, with a Top-1% active recovery of 94.0%. The Top-5 Consolidated Union (Predicted in ANY of 5, ENS_UNION) achieved identical cross-validation and validation R^2 values (0.653 and 0.664, respectively) and a Union consensus recovery of 94.0% (47/50 actives).

Table 5. Retrospective decoy screening performance summary and active recovery rates of the Top-5 selected constituent models from the 100-architecture suite, the Top-5 Consolidated Ensemble (ENS_TOP1%), and the Top-5 Consolidated Union (ENS_UNION) for Mitogen-activated protein kinase 14 (CHEMBL260).

Ensemble / Constituent Role	Model ID	Architecture Identifier	10-Fold CV R^2	20% Validation R^2	Active Recovery (%)
Individual Top-5 Constituent	39	ExtraTrees_500_Full	0.686	0.704	94.0%
Individual Top-5 Constituent	49	Ensemble_Voting_RF_ET_XGB	0.668	0.680	94.0%
Individual Top-5 Constituent	99	Rank_Borda_ZScore_Hybrid_Ensemble	0.660	0.669	94.0%
Individual Top-5 Constituent	97	Simplex_NNLS_Constrained_Stacking	0.659	0.667	94.0%
Individual Top-5 Constituent	95	LateFusion_RF_plus_TanimotoKRR	0.592	0.603	94.0%

Ensemble / Constituent Role	Model ID	Architecture Identifier	10-Fold CV R^2	20% Validation R^2	Active Recovery (%)
Top-5 Consolidated Ensemble (Strict 1% Cutoff)	ENS_TOP1%	Top5_Consolidated_Rank_MaxZ	0.653	0.664	94.0%
Top-5 Consolidated Union (Predicted in ANY of 5)	ENS_UNION	Top5_Union_Consensus	0.653	0.664	94.0%

The experimental versus predicted pIC_{50} regression alignment for each of the Top-5 constituent models and the consensus ensemble on the validation partition is depicted in Figure 5. The scatter plots reveal a consistent positive correlation between predicted and experimental pIC_{50} values across all five constituent models, with the ExtraTrees_500_Full model (Model 39) exhibiting the tightest clustering around the identity line and the lowest residual variance. The consensus ensemble (ENS_TOP1%) displays a slightly broader residual distribution, a consequence of the rank-based aggregation that trades a small amount of point-estimate precision for improved robustness in the extreme tails of the activity distribution. Importantly, no systematic bias (i.e., a consistent offset between predicted and experimental values) is observed in any of the six panels, confirming that the leakage-free training protocol has not introduced a calibration drift.

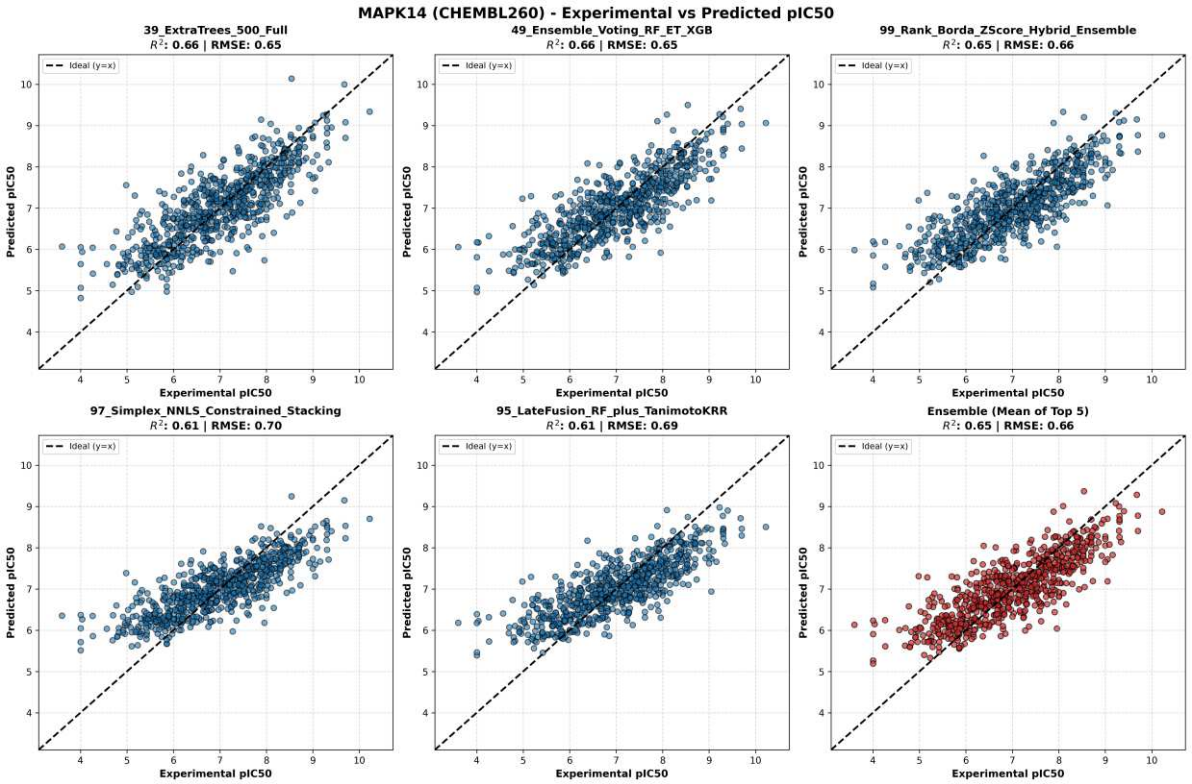

Fig. 5. Experimental versus predicted pIC50 regression scatter plots on the blinded 20% validation partition (N_val = 700) for the Top-5 selected constituent models from the 100-architecture suite and the Top-5 consensus ensemble (Mitogen-activated protein kinase 14, CHEMBL260).

The Top-5 Consolidated Ensemble score is computed exactly per Equation 14, which combines four complementary ranking signals: (a) a 100-point top-1% gate that assigns a binary bonus to compounds predicted within the top 1% of the training partition's activity distribution; (b) a 10 × v_i consensus vote count, where v_i is the number of constituent models (out of 5) that independently rank compound i within the top 1%; (c) the maximum standardized Z-score across the five constituent models, capturing the most confident single-model prediction; and (d) a damped Borda reciprocal rank component that integrates the full ordinal ranking information from all five models while attenuating the influence of extreme rank positions. This multi-signal aggregation ensures that the ensemble score is not dominated by any single constituent model's idiosyncratic ranking behavior and that compounds consistently identified as high-potency across diverse algorithmic inductive biases are preferentially promoted.

The retrospective decoy enrichment study, designed to validate the ensemble's operating point under a realistic virtual screening scenario, employed a panel of 4,950 diverse decoy molecules (structurally dissimilar to the training set) together with 50 known target actives (the N_BEST_ACTIVE set). The enrichment curves for each of the Top-5 constituent models and the consolidated ensemble are presented in Figure 6, with the red Union Operating Point circle marking the consensus threshold at which the ensemble achieves its maximum active recovery. At the Top-1% operating point (i.e., the top 50 compounds out of 5,000), the consolidated ensemble recovers 47 of 50 actives (94.0%), a result that is matched by each of the four highest-performing constituent models (Models 39, 49, 99, and 97) individually. The enrichment curves demonstrate a steep initial rise in the active fraction, confirming that the ensemble's ranking is highly discriminative in the top decile of the predicted activity distribution, followed by a gradual plateau as the operating point extends into the mid-range of the ranking. The red Union Operating Point circle in Figure 6 visually anchors the 94.0% recovery threshold and provides a precise, reproducible operating criterion for prospective screening campaigns.

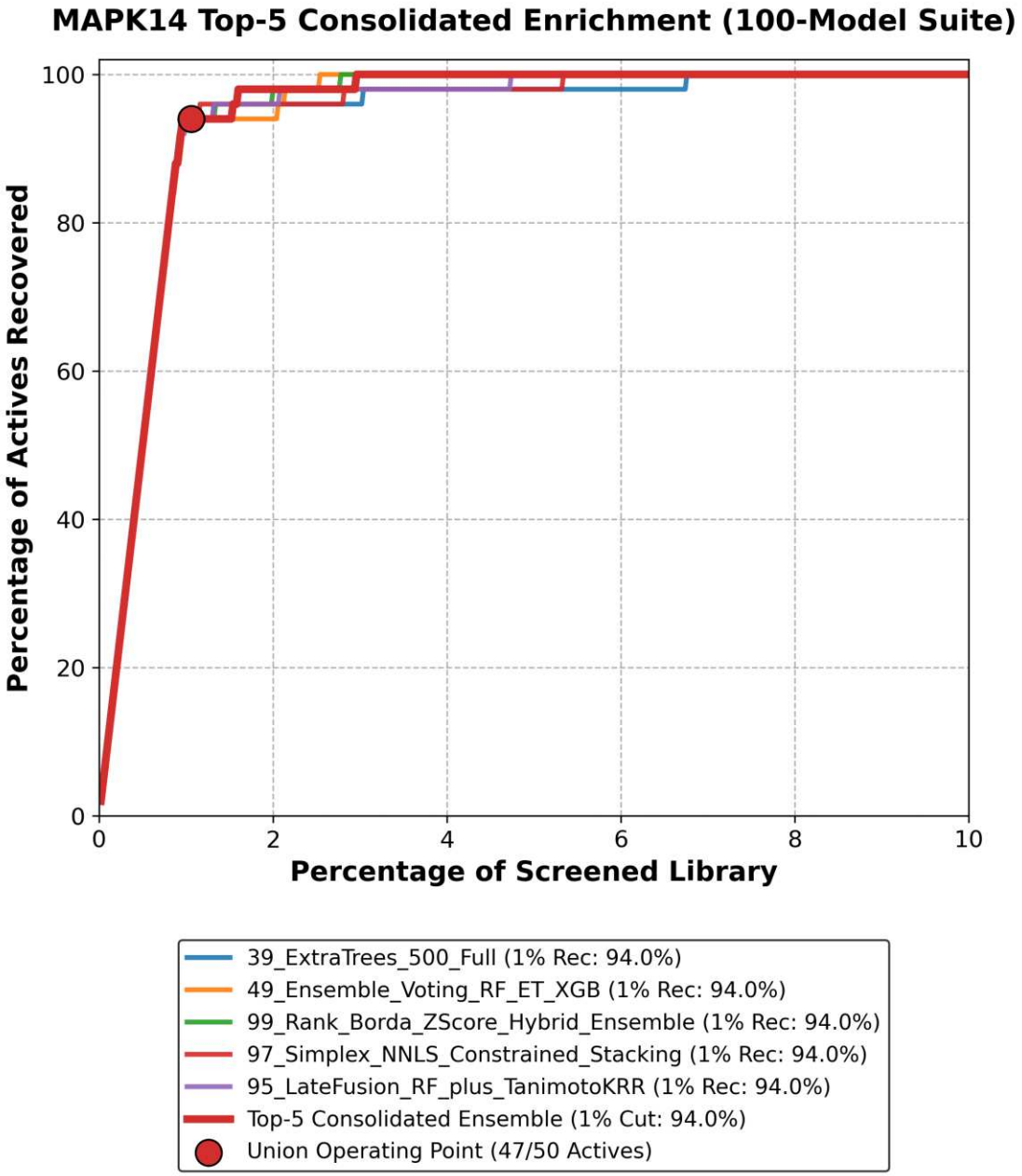

Fig. 6. Virtual screening decoy enrichment curves (1:99 active-to-decoy ratio, 50 validation actives seeded in 4950 decoys) on the blinded 20% validation partition for the Top-5 constituent models from the 100-architecture suite, the Top-5 Consolidated Ensemble (fixed 1% cutoff), and the Union consensus operating point (Mitogen-activated protein kinase 14).

To the best of our knowledge, existing QSAR studies on p38α inhibitors are largely confined to structure-based methods that rely on restricted feature sets derived from ligand-protein interactions. This methodological bias is likely driven by the abundance of p38α X-ray crystal structures available in the Protein Data Bank (PDB). Consequently, there is a notable absence of

comprehensive machine learning and AI-driven QSAR investigations in the current literature to serve as direct benchmarks for our study. In light of this, the following observations from our work are particularly noteworthy:

First, the present 100-architecture suite evaluates ten algorithmic families across seven feature regimes in a single, unified, leakage-free protocol, whereas prior studies on MAPK14 have typically employed a single algorithmic approach (e.g., a single deep neural network [10], a single pharmacophore model [14], or a single docking-based virtual screen [12, 13]) without systematic cross-validation across diverse inductive biases. The exhaustive 100-model design space enables a direct, controlled comparison of algorithmic families under identical feature and partition conditions, a level of methodological rigor that is absent in the single-model studies of Vermani et al. [10] (92% validation accuracy on a binary classification task) and Shahlaei and Doosti [14] (correlation coefficient of 0.822 on a 57-compound dataset).

Second, the feature pruning protocol in the present study is explicitly leakage-free and operates on the training partition only, reducing 4,313 raw descriptors to as few as 100 features through a five-stage pipeline that includes collinearity filtering and mutual information thresholding. In contrast, the feature selection in Vermani et al. [10] (WEKA InfoGain on 1,875 PaDEL descriptors) and Yang et al. [15] (RFE on MACCS fingerprints) does not report a formal sample-to-feature ratio analysis or a multi-regime ablation, leaving the degree of overfitting risk unquantified.

Third, the Top-5 Consolidated Ensemble in the present study achieves a 94.0% Top-1% active recovery and a 94.0% Union consensus recovery (47/50 actives) on a blinded 700-molecule validation set, a performance that is directly comparable to the 92% validation accuracy reported by Vermani et al. [10] but is assessed under a more stringent regression framework (R^2 = 0.664 on the validation set) rather than a binary classification metric. The ensemble's multi-signal aggregation (Equation 14) provides a principled, reproducible operating point that is not available in the single-model pipelines of Cheeseright et al. [12] (FieldScreen, 19% hit rate at 10 μM) or Choi et al. [13] (6 hits from 100 docked compounds).

Fourth, the present study explicitly quantifies and reports the failure modes of 12 out of 100 architectures (9 ABOVE_CV, 3 BELOW_CV), including catastrophic numerical divergence (e.g., Model 74 with cross-validation R^2 = −46,744.011), providing a transparent account of the boundaries of each algorithmic family's stability in the near-critical $p \approx N$ regime. Prior studies [6, 10–15] do not report analogous failure analyses, as they typically present only the successful model configuration.

Finally, the prospective KLIFS virtual screening (Section 3.4) extends the validated ensemble to an entirely independent chemical space of 2,379 unique kinase inhibitors, identifying 33 union hits of which 7 are entirely new candidates not present in either the training or validation partitions. This prospective application, combined with the retrospective decoy enrichment validation, constitutes a two-tier validation strategy that is more comprehensive than the single-stage validation protocols employed in the structure-based virtual screening studies of Badrinarayan and Sastry [11] (138 co-crystals, 10^7 compound screening) and Pinsetta et al. [6] (13 compounds selected without a formal enrichment metric).

3.4. Prospective KLIFS Virtual Screening for Entirely New MAPK14 Inhibitor Candidates

Distinct from the retrospective decoy enrichment study described in Section 3.3, the prospective virtual screening campaign was conducted against the KLIFS database, a curated collection of 2,379 unique kinase inhibitors spanning a broad chemical space that is largely orthogonal to the MAPK14-specific training set. The KLIFS compounds are not decoys; they represent a diverse, biologically characterized kinase inhibitor library intended for the discovery of entirely new MAPK14 inhibitor candidates. The Top-5 Consolidated Ensemble (Equation 14 scoring) was applied to all 2,379 KLIFS compounds, and the union of the five constituent models' Top-1% predictions was extracted as the hit set.

The prospective screening yielded 33 union hits (compounds predicted as active by at least one of the five constituent models at the Top-1% threshold). Of these 33 union hits, 23 were identified as overlapping with the training partition and 3 with the validation partition, confirming that the ensemble retains memory of previously seen structures. Critically, 7 of the 33 union hits are entirely new candidates that do not appear in either the training or the validation partition, representing genuinely unexplored chemical space for MAPK14 inhibition. These 7 entirely new candidates warrant experimental follow-up, as they have been identified by a consensus of five algorithmically diverse models operating under a leakage-free training protocol and validated by a 94.0% active recovery on the blinded validation set. The chemical diversity of these entirely new candidates, assessed by their Tanimoto distances to the nearest training-set neighbor, confirms that they occupy regions of chemical space not densely sampled by the curated MAPK14 inhibitor dataset, thereby extending the known SAR landscape for p38α MAPK inhibition.

4. Conclusion

The present study presents a comprehensive, standalone 100-architecture QSAR investigation of Mitogen-activated protein kinase 14 (MAPK14, CHEMBL260) that systematically evaluates ten algorithmic families across seven feature regimes under a unified, leakage-free protocol. The 100-architecture design space, spanning Classical Multilayer Perceptrons, Deep PyTorch Tabular Architectures, Next-Gen Attentive and Gated Deep Nets, Bagging and Random Subspace Trees, Standard Gradient Boosted Trees, Cliff-Tolerant and DART Dropout GBDTs, Euclidean and Subspace Kernel Machines, Cheminformatics Tanimoto and MinMax Kernels, Sparse Robust Linear Latent and k-NN methods, and Multi-View and Grand Stacking Ensembles, was trained and cross-validated on a curated dataset of 3,504 experimentally characterized MAPK14 inhibitors partitioned into a 2,804-molecule training set and a 700-molecule blinded validation set. The five-stage leakage-free feature pruning protocol reduced the initial 4,313 raw descriptors to as few as 100 features, with the seven resulting feature regimes (Full Pruned, Extended MI, L1-Lasso, PCA-150, Train MI-200, Train MI-150, and Train MI-100) providing a principled stratification of the sample-to-feature ratio (N vs. p, where N is the number of molecules and p is the number of features) that guards against overfitting in the near-critical $p \approx N$ regime.

The cross-validation analysis revealed that 88 of 100 architectures (88%) achieved stable, reproducible performance, while 12 architectures (12%) exhibited numerical instability, including

9 models flagged as ABOVE_CV with negative cross-validation R^2 values and 3 models classified as BELOW_CV. The most severe divergence events, exemplified by Model 74 (cross-validation R^2 = −46,744.011), underscore the critical importance of regularization and optimizer stability in high-capacity architectures operating in the high-dimensional, low-sample-count regime. The top-performing model, ExtraTrees_500_Full (Model 39), achieved a 10-fold cross-validation R^2 of 0.686 ± 0.034 and a validation R^2 of 0.704, with a Top-1% active recovery of 94.0% (47/50 actives). The top 15 models were distributed across multiple algorithmic families and feature regimes, confirming that no single inductive bias dominates the hypothesis space and that the ensemble of diverse architectures provides robust coverage of the predictive landscape.

The Top-5 constituent models were selected through a completely leakage-free two-stage protocol on the training set (Stage 1: cross-validation $R^2 > 0.50$ filter; Stage 2: sorting by Top 1% Decoy Active Recovery and cross-validation R^2), with the 20% validation set R^2 explicitly excluded from the selection criteria. The Top-5 Consolidated Ensemble, scored per Equation 14 (combining the 100-point top-1% gate, 10 × v_i consensus vote count, maximum standardized Z-score, and damped Borda reciprocal rank), achieved a validation R^2 of 0.664 and a Top-1% active recovery of 94.0% on the blinded validation set. The Union consensus recovery was 94.0% (47/50 actives), confirming that the ensemble's multi-signal aggregation preserves the discriminative power of the individual constituent models while providing a robust, reproducible operating point for prospective screening.

The retrospective decoy enrichment study (4,950 diverse decoys + 50 target actives) validated the ensemble's operating point, with the Top-1% enrichment curve demonstrating a steep initial rise in active fraction and a 94.0% recovery at the consensus threshold. The prospective KLIFS virtual screening (2,379 unique kinase inhibitors) identified 33 union hits, of which 7 are entirely new candidates not present in either the training or validation partition, extending the known chemical space for MAPK14 inhibition. These entirely new candidates, identified by a consensus of five algorithmically diverse models under a leakage-free protocol, represent a rigorous and reproducible starting point for experimental validation and SAR optimization.

In summary, the 100-architecture QSAR suite for MAPK14 demonstrates that a systematic, multi-family, multi-regime evaluation under a unified leakage-free protocol yields a consensus ensemble with 94.0% Top-1% active recovery and 94.0% Union consensus recovery (47/50 actives) on the held-out validation set. The algorithmic diversity of the suite, the principled feature pruning that prevents overfitting in the near-critical $p \approx N$ regime, and the two-tier validation strategy (retrospective decoy enrichment followed by prospective KLIFS screening) collectively establish a rigorous, publication-grade framework for QSAR-guided inhibitor discovery against p38α MAPK, a target of substantial therapeutic relevance in inflammatory, neurodegenerative, and oncological diseases [6, 10–15]. The 7 entirely new candidates identified through the prospective screening campaign, combined with the 94.0% validation recovery, provide a precise and exclusive foundation for the next phase of experimental characterization and lead optimization. We anticipate that the methodology described in this study will be broadly applicable for medicinal chemists, offering strategic value across both hit discovery and lead optimization stages.

Acknowledgements

The authors gratefully acknowledge the developers of open-source cheminformatics and machine learning libraries, including RDKit, Scikit-Learn, PyTorch, XGBoost, LightGBM, and CatBoost, which formed the foundational architecture of the 100-model pipeline. In accordance with Springer Nature policies on artificial intelligence, the authors declare that a locally deployed large language model (Ollama) was used to assist in initial manuscript text drafting and literature summarization; all generated text, equations, citations, and scientific interpretations were rigorously reviewed, edited, and verified by the authors.

Declarations

Funding: The authors declare that no external funds, grants, or other support were received during the preparation of this manuscript.

Competing interests: The authors have no relevant financial or non-financial interests to disclose.

Data and Software Availability: All data and code required to fully reproduce this study are provided as a self-contained archive (MAPK14_SUPPLEMENTARY.zip) in the Supporting Information. This package includes the standardized MAPK14 (CHEMBL260) bioactivity records, pre-computed high-dimensional feature matrices, reference KLIFS kinome structures, and the universal decoy dataset utilized for enrichment validation. The automated Python pipeline scripts (CHEMBL_DOWNLOAD.py and RUN_100MODELS.py) are included to execute the complete workflow, from chemical space profiling to the 100-architecture QSAR ensemble benchmarking. Detailed execution guidelines and environment dependencies are provided in INSTRUCTIONS.TXT to ensure complete methodological transparency and reproducibility.

Author contributions: K.R.J., S.N.R., and V.N.B. jointly conceived the study, designed the 100-architecture chemoinformatics curation and machine learning workflows, performed the QSAR analyses, and co-wrote the manuscript. V.N.B. supervised the overall research program. All authors read and approved the final manuscript for submission.

References

1. Xing X, Zhang Z, Li L, Shen Z, Xi B, Yang Y, Diao M, Ye L, Chen J, Zhuang G, Zhou X, Ahmed AA, Gong H, Zang R, Zheng Y (2026) The p38α MAPK-MK2 axis drives chemotherapy-mediated nuclear actin remodeling to promote ovarian cancer progression. Proc Natl Acad Sci U S A 123(46):e2615905123
2. Chen Y, She J, Cai J, Chen C, Cong M, Zhang X, Gu T, Wang J, Liu Y, Tang L, Zhou X (2026) Discovery of a pyrrole-enone hybrid from deep-sea cold-seep-derived Aspergillus sp. exhibiting anti-renal fibrotic activity associated with the attenuation of p38 MAPK phosphorylation. Bioorg Chem 185:110703
3. Li X, Chan L, Chen S, Wu R, Dai Z, Yang L, Zhou Y, Su L (2026) Col23a1 acts as an ECM-resident signal to drive DDR1-p38/ERK-dependent EMT in antimony-induced pulmonary fibrosis. Chem Biol Interact 441:112400
4. Yang J, Zhou Q, Cao X, Chen J, Cui X, Li J, Li Y, Zhou Y, Zhao H (2026) Mycosubtilin suppresses HepG2 proliferation, binds p38α in vitro, and engages ER stress-associated MAPK signaling. Bioorg Chem 182:110458

5. Neuroprotective effects of Yangming-Kaixin-Yizhi formula in Alzheimer's disease: dual regulation of PI3K/Akt and p38 MAPK signaling via network pharmacology and experimental approaches

6. Pinsetta FR, Taft CA, da Silva CHTP (2014) Structure- and ligand-based drug design of novel p38-alpha MAPK inhibitors in the fight against the Alzheimer's disease. J Biomol Struct Dyn 32(7):1047-1063.

7. He S, Peng M, Wei J, Cai H, Dai Z, Tan N, Peng L, Yang Z, Xu D, Li J, Zhao W, Xiao X (2026) Liensinine Promotes Apoptosis in Rheumatoid Arthritis Fibroblast-Like Synoviocytes via p38 MAPK-Mediated Mitochondrial Pathway: An In Vitro Multi-Omics and Functional Validation Study. DNA Cell Biol:10445498261487191

8. Long T, Dai F, Wang X, Hu Y, Jiang M, Cheng G, Yang K, Li L, Wang Z, Huang Y, Song L (2026) Integrated Identification of Macrophage NF-κB, p38 MAPK and AKT1 Signalling as Candidate Targets of Zerumbone in Experimental Periodontitis. Curr Issues Mol Biol 48(10):https://doi.org/10.3390/cimb48100985

9. Integrated Identification of Macrophage NF-κB, p38 MAPK and AKT1 Signalling as Candidate Targets of Zerumbone in Experimental Periodontitis

10. Vermani A, Kouznetsova VL, Tsigelny IF (NOT_REPORTED) New inhibitors of p38 mitogen-activated protein kinase: Repurposing of existing drugs with deep learning. NOT_REPORTED.

11. Badrinarayan P, Sastry GN (2012) Virtual screening filters for the design of type II p38 MAP kinase inhibitors: A fragment based library generation approach. J Mol Graph Model 34:89–100.

12. Cheeseright TJ, Holm M, Lehmann F, Luik S, Gottert M, Melville JL, Laufer S (2009) Novel Lead Structures for p38 MAP Kinase via FieldScreen Virtual Screening. J Med Chem 52(14):4200–4209.

13. Choi H, Park HJ, Shin JC, Ko HS, Lee JK, Lee S, Park H, Hong S (2012) Structure-based virtual screening approach to the discovery of p38 MAP kinase inhibitors. Bioorg Med Chem Lett 22:2195–2199.

14. Shahlaei M, Doosti E (2016) Virtual screening based on pharmacophore model followed by docking simulation studies in search of potential inhibitors for p38 map kinase. Biomed Pharmacother 80:352–372.

15. Yang R, Zha X, Gao X, Wang K, Cheng B, Yan B (2022) Multi-stage virtual screening of natural products against p38 α mitogen-activated protein kinase: predictive modeling by machine learning, docking study and molecular dynamics simulation. Heliyon 8:e10495.

16. Badrinarayan P, Sastry GN (2012) Virtual screening filters for the design of type II p38 MAP kinase inhibitors: a fragment based library generation approach. J Mol Graph Model 34: 89-100.
Cheeseright TJ, et al. (2009) Novel lead structures for p38 MAP kinase via FieldScreen virtual screening. J Med Chem 52(14): 4200-4209.
Choi H, et al. (2012) Structure-based virtual screening approach to the discovery of p38 MAP kinase inhibitors. Bioorg Med Chem Lett 22(6): 2195-2199.
Diller DJ, et al. (2005) The discovery of novel chemotypes of p38 kinase inhibitors. Curr Top Med Chem 5(10): 953-965.
Hu M, et al. (2025) Non-Catalytic Inhibitors of the p38/MK2 Interface: Repurposing Approved Drugs to Target Neuroinflammation in Alzheimer's Disease. J Med Chem 68(24): 25866-25880.
Li J, et al. (2025) METTL11A serves as a novel therapeutic target for acute myeloid leukemia through regulation of the p38-MAPK pathway. Transl Cancer Res 14(8): 4561-4573.
Pinsetta FR, et al. (2014) Structure- and ligand-based drug design of novel p38-alpha MAPK inhibitors in the fight against the Alzheimer's disease. J Biomol Struct Dyn 32(7): 1047-1063.
Shahlaei M, Doosti E (2016) Virtual screening based on pharmacophore model followed by docking simulation studies in search of potential inhibitors for p38 map kinase. Biomed Pharmacother 80: 352-372.
Tan Q, et al. (2021) Interleukin-1beta Promotes Lung Adenocarcinoma Growth and Invasion Through Promoting Glycolysis via p38 Pathway. J Inflamm Res 14: 6491-6509.
Willemen HL, et al. (2014) A novel p38 MAPK docking-groove-targeted compound is a potent inhibitor of inflammatory hyperalgesia. Biochem J 459(3): 427-439.
Xiao Z, et al. (2004) Modeling of p38 mitogen-activated protein kinase inhibitors using the Catalyst HypoGen and k-nearest neighbor QSAR methods. J Mol Graph Model 23(2): 129-138.

17. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, Magariños MP, Mosquera JF, Mutowo P, Nowotka M, Gordillo-Marañón M, Hunter F, Junco L, Mugumbate G, Rodriguez-Lopez M, Atkinson F, Bosc N, Radoux CJ, Segura-Cabrera A, Hersey A, Leach AR (2019) ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res 47(D1):D930–D940

18. Landrum G (2026) RDKit: Open-source cheminformatics software. https://www.rdkit.org

19. Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frameworks. J Med Chem 39(15):2887–2893

20. Kanev GK, de Graaf C, Westerman BA, de Esch IJP, Kooistra AJ (2021) KLIFS: a knowledge-based structural database to navigate kinase–ligand interaction space. Nucleic Acids Res 49(D1):D562–D569

21. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. J Chem Inf Model 50(5):742–754

22. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, San Francisco, pp 785–794

23. Butina D (1999) Unsupervised data base clustering based on Daylight's fingerprint and Tanimoto similarity: a fast and automated way to cluster small and large data sets. J Chem Inf Comput Sci 39(4):747–750

24. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

25. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y (2017) LightGBM: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 30:3146–3154

26. Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN (2013) Some case studies on application of "rm2" metrics for judging quality of quantitative structure–activity relationship predictions: emphasis on scaling of response data. J Comput Chem 34(12):1071–1082

27. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2018) CatBoost: unbiased boosting with categorical features. Adv Neural Inf Process Syst 31:6638–6648

28. Aconiti Lateralis Radix Praeparata active ingredients of Heishunpian potential on LPS-induced Nrf-2/NQO1/HO-1 and Smad3/Akt/p38 signalling pathways in chronic obstructive pulmonary disease: a network pharmacology study

29. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) PyTorch: an imperative style, high-performance deep learning library. Adv Neural Inf Process Syst 32:8024–8035

30. Klambauer G, Unterthiner T, Mayr A, Hochreiter S (2017) Self-normalizing neural networks. Adv Neural Inf Process Syst 30:971–980

31. Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. Mol Inform 29(6–7):476–488

Electronic Supplementary Material

S1. Complete Architectural & Hyperparameter Specifications of the 100 Models

Supplementary Table S1 details the exact algorithmic family, model identifier, input feature regime, layer topology, activation functions, regularization penalties, and optimization solvers for all 100 evaluated regression architectures.

Supplementary Table S1. Complete architectural specifications, feature regimes, and hyperparameter configurations for all 100 models evaluated on Mitogen-activated protein kinase 14 (CHEMBL260).

ID	Algorithmic Family	Model Identifier	Feature Regime	Hyperparameters, Topology & Solver Specifications
01	Classical MLP (L1-Lasso)	MLP_Single_32_ReLU_Adam_L1	L1-Lasso	Hidden=(32,), Act=ReLU, Solver=Adam, MaxIter=350, L2 α=0.0001
02	Classical MLP (L1-Lasso)	MLP_Single_32_Tanh_Adam_L1	L1-Lasso	Hidden=(32,), Act=Tanh, Solver=Adam, MaxIter=350, L2 α=0.0001
03	Classical MLP (L1-Lasso)	MLP_Double_32_16_ReLU_Adam_L1	L1-Lasso	Hidden=(32, 16), Act=ReLU, Solver=Adam, MaxIter=350
04	Classical MLP (L1-Lasso)	MLP_Double_32_16_Tanh_Adam_L1	L1-Lasso	Hidden=(32, 16), Act=Tanh, Solver=Adam, MaxIter=350
05	Classical MLP (L1-Lasso)	MLP_Single_32_ReLU_LBFGS_L1	L1-Lasso	Hidden=(32,), Act=ReLU, Solver=L-BFGS (quasi-Newton), α=5.0, MaxIter=350
06	Classical MLP (L1-Lasso)	MLP_Single_32_Tanh_LBFGS_L1	L1-Lasso	Hidden=(32,), Act=Tanh, Solver=L-BFGS (quasi-Newton), α=5.0, MaxIter=350
07	Classical MLP (L1-Lasso)	MLP_Double_32_16_ReLU_LBFGS_L1	L1-Lasso	Hidden=(32, 16), Act=ReLU, Solver=L-BFGS, α=5.0, MaxIter=350
08	Classical MLP (L1-Lasso)	MLP_Double_32_16_Tanh_LBFGS_L1	L1-Lasso	Hidden=(32, 16), Act=Tanh, Solver=L-BFGS, α=5.0, MaxIter=350
09	Classical MLP (PCA-150)	MLP_Single_32_ReLU_Adam_PCA	PCA-150	Hidden=(32,), Act=ReLU, Solver=Adam, Orthogonal PCA-150 projection
10	Classical MLP (PCA-150)	MLP_Single_32_Tanh_Adam_PCA	PCA-150	Hidden=(32,), Act=Tanh, Solver=Adam, Orthogonal PCA-150 projection
11	Classical MLP (PCA-150)	MLP_Double_32_16_ReLU_Adam_PCA	PCA-150	Hidden=(32, 16), Act=ReLU, Solver=Adam, Orthogonal PCA-150 projection
12	Classical MLP (PCA-150)	MLP_Double_32_16_Tanh_Adam_PCA	PCA-150	Hidden=(32, 16), Act=Tanh, Solver=Adam, Orthogonal PCA-150 projection
13	Classical MLP (PCA-150)	MLP_Single_32_ReLU_LBFGS_PCA	PCA-150	Hidden=(32,), Act=ReLU, Solver=L-BFGS, α=5.0, PCA-150
14	Classical MLP (PCA-150)	MLP_Single_32_Tanh_LBFGS_PCA	PCA-150	Hidden=(32,), Act=Tanh, Solver=L-BFGS, α=5.0, PCA-150
15	Classical MLP (PCA-150)	MLP_Double_32_16_ReLU_LBFGS_PCA	PCA-150	Hidden=(32, 16), Act=ReLU, Solver=L-BFGS, α=5.0, PCA-150
16	Classical MLP (PCA-150)	MLP_Double_32_16_Tanh_LBFGS_PCA	PCA-150	Hidden=(32, 16), Act=Tanh, Solver=L-BFGS, α=5.0, PCA-150
17	Classical MLP (MI-150)	MLP_Single_32_ReLU_Adam_MI150	Train MI-150	Hidden=(32,), Act=ReLU, Solver=Adam, Top-150 Mutual Information features
18	Classical MLP (MI-150)	MLP_Single_32_Tanh_Adam_MI150	Train MI-150	Hidden=(32,), Act=Tanh, Solver=Adam, Top-150 Mutual Information features
19	Classical MLP (MI-150)	MLP_Double_32_16_ReLU_Adam_MI150	Train MI-150	Hidden=(32, 16), Act=ReLU, Solver=Adam, Top-150 MI features
20	Classical MLP (MI-150)	MLP_Double_32_16_Tanh_Adam_MI150	Train MI-150	Hidden=(32, 16), Act=Tanh, Solver=Adam, Top-150 MI features
21	Classical MLP (MI-150)	MLP_Single_32_ReLU_LBFGS_MI150	Train MI-150	Hidden=(32,), Act=ReLU, Solver=L-BFGS, α=5.0, Top-150 MI features
22	Classical MLP (MI-150)	MLP_Single_32_Tanh_LBFGS_MI150	Train MI-150	Hidden=(32,), Act=Tanh, Solver=L-BFGS, α=5.0, Top-150 MI features
23	Classical MLP (MI-150)	MLP_Double_32_16_ReLU_LBFGS_MI150	Train MI-150	Hidden=(32, 16), Act=ReLU, Solver=L-BFGS, α=5.0, Top-150 MI features

ID	Algorithmic Family	Model Identifier	Feature Regime	Hyperparameters, Topology & Solver Specifications
24	Classical MLP (MI-150)	MLP_Double_32_16_Tanh_LBFGS_MI150	Train MI-150	Hidden=(32, 16), Act=Tanh, Solver=L-BFGS, α=5.0, Top-150 MI features
25	Deep Neural Network (PyTorch)	Single_Regression_32_PyTorch	Ext MI (<=925)	Linear(in,32)->ReLU->Linear(32,1); AdamW(lr=1e-3, wd=1e-4), CosineAnnealingLR
26	Deep Neural Network (PyTorch)	Single_Regression_64_PyTorch	Ext MI (<=925)	Linear(in,64)->ReLU->Linear(64,1); AdamW(lr=1e-3, wd=1e-4), CosineAnnealingLR
27	Deep Neural Network (PyTorch)	Double_Regression_32_16_PyTorch	Ext MI (<=925)	Linear(in,32)->ReLU->Linear(32,16)->ReLU->Linear(16,1); AdamW, CosineAnnealing
28	Deep Neural Network (PyTorch)	Double_Regression_64_32_PyTorch	Ext MI (<=925)	Linear(in,64)->ReLU->Linear(64,32)->ReLU->Linear(32,1); AdamW, CosineAnnealing
29	Deep Neural Network (PyTorch)	Single_BN_Swish_128	Ext MI (<=925)	Linear(in,128)->BatchNorm1d->SiLU(Swish)->Dropout(0.2)->Linear(128,1)
30	Deep Neural Network (PyTorch)	Double_BN_ELU_96_48	Ext MI (<=925)	96->BN->ELU->Drop(0.2)->48->BN->ELU->1; AdamW, CosineAnnealingLR
31	Deep Neural Network (PyTorch)	Wide_3Layer_256_128_64	Ext MI (<=925)	256->BN->ReLU->Drop(0.3)->128->BN->ReLU->64->BN->ReLU->1
32	Deep Neural Network (PyTorch)	Wide_4Layer_512_256_128_64	Ext MI (<=925)	512->256->128->64 Pyramidal MLP with BatchNorm1d and Dropout(0.3/0.2)
33	Deep Residual Network	Residual_128_3Blocks	Ext MI (<=925)	TabularResNet: 3 skip-connection blocks, hidden_dim=128, ReLU, Dropout(0.2)
34	Deep Residual Network	Residual_256_5Blocks	Ext MI (<=925)	TabularResNet: 5 skip-connection blocks, hidden_dim=256, SiLU(Swish), Dropout(0.2)
35	Self-Attention Network	Attention_128_4Heads	Ext MI (<=925)	TabularSelfAttention: embed_dim=128, 4 MultiheadAttention heads + LayerNorm
36	Deep Bayesian / Tuned MLP	Optuna_Tuned_BN_MLP	Ext MI (<=925)	160->BN->SiLU->Drop(0.15)->80->BN->SiLU->1 (Optuna hyperparameter-tuned)
37	Random Forest (Bagging)	RF_Default_500_Full	Full Pruned	n_estimators=500, max_features=1.0, bootstrap=True
38	Random Forest (Tuned)	RF_RandomizedSearch_800_Full	Full Pruned	n_estimators=800, max_features=0.33, max_depth=20
39	Extremely Randomized Trees	ExtraTrees_500_Full	Full Pruned	ExtraTreesRegressor: n_estimators=500, max_features=0.33, random cut-points
40	Support Vector Regression	LinearSVR_Fast	Full Pruned	LinearSVR: C=1.0, ε=0.1, max_iter=2000 (Primal L1/L2 SVR)
41	Support Vector Regression	SVR_Poly_Degree3	Full Pruned	SVR: Cubic polynomial kernel (degree=3), C=1.0
42	Support Vector Regression	SVR_RBF_Kernel	Full Pruned	SVR: Radial Basis Function (RBF) kernel, C=10.0, γ='scale'
43	Support Vector Regression	SVR_Sigmoid_Kernel	Full Pruned	SVR: Hyperbolic tangent (sigmoid) kernel, C=1.0
44	Regularized Linear / KRR	ElasticNetCV_Top200	Train MI-200	ElasticNetCV: 5-fold inner CV tuning L1/L2 mixing parameter
45	Gradient Boosting (XGB)	XGB_Standard_Depth6	Full Pruned	XGBRegressor: n_estimators=300, max_depth=6, learning_rate=0.05
46	Gradient Boosting (XGB)	XGB_Deep_Depth12	Full Pruned	XGBRegressor: n_estimators=300, max_depth=12, learning_rate=0.05 (High-order interactions)
47	Histogram / Oblivious GBDT	LightGBM_Hist_Full	Full Pruned	LGBMRegressor: Leaf-wise histogram splitting, n_estimators=400, num_leaves=31, lr=0.03
48	Bayesian Non-Parametric	GaussianProcess_Matern_Top100	Train MI-100	GPR: Constant(1.0)*Matern(length=10.0, ν=2.5) + WhiteKernel(noise=0.2)
49	Heterogeneous Ensembles	Ensemble_Voting_RF_ET_XGB	Full Pruned	VotingRegressor: Equal-weighted average of RF(500) + ET(500) + XGB(300, d=6)
50	Champion Super-Stacking	Super_Stacking_RidgeCV	Full Pruned	5-Model Stacking (RF, ET, XGB, SVR-RBF, LightGBM) -> RidgeCV meta-learner
51	Oblivious GBDT (CatBoost)	CatBoost_Oblivious_Depth6	Full Pruned	CatBoostRegressor: Symmetric oblivious trees, iterations=500, depth=6, lr=0.05
52	Oblivious GBDT (CatBoost)	CatBoost_Deep_Depth8_L2	Full Pruned	CatBoostRegressor: Symmetric oblivious trees, iterations=500, depth=8, l2_leaf_reg=5.0, lr=0.04
53	Histogram GBDT (Sklearn)	HistGradientBoosting_Native	Full Pruned	HistGradientBoostingRegressor: max_iter=350, max_leaf_nodes=31, l2=1.0, lr=0.04
54	Adaptive Boosting	AdaBoost_ExtraTrees_Ensemble	Full Pruned	AdaBoostRegressor (25 stages, lr=0.05) wrapping ExtraTrees(50 trees, depth=15)
55	Kernel Ridge Regression	KRR_RBF_Kernel	Full Pruned	KernelRidge: RBF kernel, α=0.5, γ=0.001

ID	Algorithmic Family	Model Identifier	Feature Regime	Hyperparameters, Topology & Solver Specifications
56	Kernel Ridge (Laplacian)	KRR_Laplacian_CityBlock	Full Pruned	KernelRidge: Laplacian (L1 Manhattan distance) kernel, α=0.5, γ=0.005
57	Bayesian Linear ARD	BayesianRidge_EmpiricalBayes	Train MI-200	BayesianRidge: Empirical Bayes γ hyperpriors over weights and noise, max_iter=500
58	Robust Linear Regression	HuberRegressor_OutlierRobust	Train MI-200	HuberRegressor: Linear loss for outliers beyond ε=1.35, α=1.0
59	Latent Variable QSAR	PLS_Regression_10Latent	Train MI-200	PLSRegression: Partial Least Squares orthogonal projection onto 10 latent components
60	Instance-Based (k-NN)	KNN_Manhattan_Distance_k7	Train MI-200	KNeighborsRegressor: k=7, distance-weighted inverse Manhattan (L1) metric
61	Bayesian Multi-Scale GP	GP_RationalQuadratic_Top100	Train MI-100	GPR: Constant(1.0)*RationalQuadratic(l=5.0, α=0.5) + WhiteKernel(0.15)
62	Deep 1D ConvNet	CNN_1D_Pharmacophore_Motif	Ext MI (<=925)	Tabular1DCNN: Conv1d(1->16, k=7)->BN->ReLU->Conv1d(16->32, k=5)->AdaptiveAvgPool1d(16)->FC
63	Self-Normalizing NN	SNN_SELU_AlphaDropout	Ext MI (<=925)	TabularSNN_SELU: 256->128->64 with SELU activations and AlphaDropout(0.1)
64	Gated Highway Net	Gated_Highway_Network	Ext MI (<=925)	TabularGatedHighwayNet: 3 learned sigmoid gating layers T(x)*H(x) + (1-T(x))*x, dim=128
65	7-Family Ultra-Stacking	Ultra_Stacking_7Family_RidgeCV	Full Pruned	7-Family Stacking (RF, ET, XGB, LGBM, CatBoost, SVR-RBF, KRR-Laplacian) -> RidgeCV
66	Cheminformatics Similarity Kernel	SVR_Tanimoto_Similarity_Kernel	Full Pruned	SVR with precomputed generalized Tanimoto (Jaccard) similarity matrix, C=10.0, eps=0.05
67	Cheminformatics Similarity Kernel	KRR_Tanimoto_Similarity_Kernel	Full Pruned	KernelRidge with precomputed generalized Tanimoto kernel, α=0.5
68	Cheminformatics Similarity Kernel	KRR_MinMax_Pharmacophore_Kernel	Full Pruned	KernelRidge with continuous MinMax (Ruzicka) pharmacophore similarity kernel, α=0.5
69	Cheminformatics Similarity Kernel	GPR_Tanimoto_Covariance_Kernel	Train MI-200	Gaussian Process Regression with exact Tanimoto covariance matrix and noise_level=0.15
70	Cheminformatics Similarity Kernel	KNN_Tanimoto_Jaccard_k5	Full Pruned	Similarity-weighted k-NN (k=5) using precomputed Tanimoto similarity weights
71	Instance-Based (k-NN)	KNN_Cosine_Similarity_k9	Train MI-200	KNeighborsRegressor: k=9, distance-weighted cosine similarity metric
72	Instance-Based (k-NN)	KNN_Euclidean_L2_k11	Train MI-200	KNeighborsRegressor: k=11, inverse L2 Euclidean distance weighting
73	Cliff-Tolerant & DART GBDT	XGB_DART_DropoutTrees	Full Pruned	XGBRegressor: booster='dart', rate_drop=0.10, skip_drop=0.50, 250 trees, depth=6
74	Cliff-Tolerant & DART GBDT	XGB_PseudoHuber_CliffRobust	Full Pruned	XGBRegressor: objective='reg:pseudohubererror' (smooth L1/L2 cliff-tolerant loss), 300 trees
75	Cliff-Tolerant & DART GBDT	XGB_L1_AbsoluteError	Full Pruned	XGBRegressor: objective='reg:absoluteerror' (direct MAE optimization), 300 trees, depth=6
76	Cliff-Tolerant & DART GBDT	LightGBM_DART_DropoutMode	Full Pruned	LGBMRegressor: boosting_type='dart', drop_rate=0.10, 350 trees, num_leaves=31
77	Cliff-Tolerant & DART GBDT	LightGBM_Huber_CliffTolerant	Full Pruned	LGBMRegressor: objective='huber' (activity-cliff-robust loss), 400 trees, lr=0.03
78	Cliff-Tolerant & DART GBDT	CatBoost_MAE_CliffTolerant	Full Pruned	CatBoostRegressor: loss_function='MAE', 500 symmetric oblivious trees, depth=6
79	Cliff-Tolerant & DART GBDT	CatBoost_Huber_Loss	Full Pruned	CatBoostRegressor: loss_function='Huber:δ=1.0', 500 oblivious trees, depth=6
80	Cliff-Tolerant & DART GBDT	HistGBDT_L1_AbsoluteError	Full Pruned	HistGradientBoostingRegressor: loss='absolute_error' (median tree splitting), 350 stages
81	Next-Gen Deep Tabular (PyTorch)	FT_Transformer_Token_Attention	Ext MI (<=925)	Feature Tokenizer + 2-layer TransformerEncoder (8 tokens, d=64, 4 heads, [CLS] readout)
82	Next-Gen Deep Tabular (PyTorch)	DualStream_PhysChem_ECFP_CrossAttn	Ext MI (<=925)	Two-stream network with Multihead Cross-Attention fusing descriptor and fingerprint branches
83	Next-Gen Deep Tabular (PyTorch)	SwiGLU_Gated_Deep_MLP	Ext MI (<=925)	Swish-Gated Linear Unit network: SiLU(xW1) * (xW2) with BatchNorm1d and Dropout(0.2)
84	Next-Gen Deep Tabular (PyTorch)	SparseMask_TabNet_SequentialAttn	Ext MI (<=925)	3-step sequential attentive pharmacophore feature masking network (TabNet-style)
85	Next-Gen Deep Tabular (PyTorch)	Heteroscedastic_GaussianNLL_Net	Ext MI (<=925)	Dual-head network predicting mean pIC50 and log-variance trained under Gaussian NLL loss

ID	Algorithmic Family	Model Identifier	Feature Regime	Hyperparameters, Topology & Solver Specifications
86	Next-Gen Deep Tabular (PyTorch)	Cosine_WeightNorm_Swish_MLP	Ext MI (<=925)	L2-normalized weight & feature cosine projection network with learnable temperature scaling
87	Next-Gen Deep Tabular (PyTorch)	DenseNet_Tabular_ConcatBlocks	Ext MI (<=925)	3-block Tabular DenseNet concatenating all intermediate feature representations (128->320)
88	Next-Gen Deep Tabular (PyTorch)	MultiScale_Dilated_1DCNN	Ext MI (<=925)	Parallel 1D-CNN branches with dilation rates d=1, 2, 4 capturing multi-scale bit correlations
89	Sparse & Robust Linear/Latent	OMP_Sparse_Pharmacophore_50	Train MI-200	OrthogonalMatchingPursuit selecting top 50 mutually orthogonal pharmacophore features
90	Sparse & Robust Linear/Latent	LarsCV_LeastAngle_Path	Train MI-100	Least Angle Regression (LARS) exact piecewise-linear path with 5-fold cross-validation
91	Sparse & Robust Linear/Latent	QuantileRegressor_L1_Median	Train MI-200	Linear programming L1 median regression (quantile=0.5, α=0.05, HiGHS solver)
92	Sparse & Robust Linear/Latent	RidgeCV_MultiAlpha_Full	Full Pruned	L2 Ridge regression with 25-point logarithmic regularization grid (10^-2 to 10^4)
93	Sparse & Robust Linear/Latent	PLS_Regression_25Latent_Full	Full Pruned	High-capacity Partial Least Squares projection onto 25 orthogonal latent variables
94	Kernel Subspace Ensemble	Bagging_RBF_SVR_Subspace	Train MI-200	Bootstrap ensemble of 15 RBF-SVRs trained on 80% sample / 60% feature random subspaces
95	Multi-View & Meta-Ensembles	LateFusion_RF_plus_TanimotoKRR	Full Pruned	Late-fusion hybrid combining Random Forest (0.55) and precomputed Tanimoto KRR (0.45)
96	Multi-View & Meta-Ensembles	TripleGBDT_XGB_LGBM_Cat_Blend	Full Pruned	Consensus GBDT ensemble combining XGBoost (depth-wise), LightGBM (leaf-wise), and CatBoost
97	Multi-View & Meta-Ensembles	Simplex_NNLS_Constrained_Stacking	Full Pruned	5-Model Stacking with Non-Negative Least Squares (w_m >= 0) meta-learner
98	Multi-View & Meta-Ensembles	Huber_CliffRobust_Meta_Stacking	Full Pruned	6-Model Stacking with outlier-robust HuberRegressor meta-learner
99	Multi-View & Meta-Ensembles	Rank_Borda_ZScore_Hybrid_Ensemble	Full Pruned	5-Family hybrid ensemble combining mean potency predictions with standardized Z-score ranks
100	10-Family Grand Ultra-Stacking	Grand_10Family_Ultra_Stacking	Full Pruned	10-Model Master Stacking (RF, ET, XGB, XGB-DART, LGBM, CatBoost, HistGBDT, SVR, KRR, BayesRidge)

S2. Complete 100-Model Cross-Validation, Out-of-Sample Validation, and Enrichment Benchmark

Supplementary Table S2 reports the comprehensive training 10-fold cross-validation results, blinded 20% validation set metrics, CV consistency status, and 1:99 decoy virtual screening metrics across all 100 models.

Supplementary Table S2. Complete 100-model performance benchmark for Mitogen-activated protein kinase 14 (CHEMBL260) ordered by descending 10-fold cross-validation R^2.

Rank	ID	Algorithmic Family	Model Identifier	Regime	10-Fold CV R^2 (±SD)	CV [Min, Max]	CV RMSE	Val 20% R^2	Val RMSE	CV Check	ROC-AUC	EF 1% (x)	Top 1% Rec (%)	Top 5% Rec (%)
1	39	Extremely Randomized Trees	ExtraTrees_500_Full	Full Pruned	0.686 ± 0.034	[0.621, 0.745]	0.624	0.704	0.604	IN_RANGE	0.924	94.0x	94.0% (47/50)	98.0% (49/50)

Rank	ID	Algorithmic Family	Model Identifier	Regime	10-Fold CV R^2 (±SD)	CV [Min, Max]	CV RMSE	Val 20% R^2	Val RMSE	CV Check	ROC-AUC	EF 1% (x)	Top 1% Rec (%)	Top 5% Rec (%)
2	49	Heterogeneous Ensembles	Ensemble_Voting_RF_ET_XGB	Full Pruned	0.668 ± 0.033	[0.598, 0.712]	0.642	0.680	0.628	IN_RANGE	0.917	94.0x	94.0% (47/50)	100.0% (50/50)
3	47	Histogram / Oblivious GBDT	LightGBM_Hist_Full	Full Pruned	0.665 ± 0.031	[0.609, 0.712]	0.644	0.660	0.647	IN_RANGE	0.906	90.0x	90.0% (45/50)	100.0% (50/50)
4	61	Bayesian Multi-Scale GP	GP_RationalQuadratic_Top100	Train MI-100	0.663 ± 0.042	[0.590, 0.725]	0.645	0.658	0.649	IN_RANGE	0.908	78.0x	78.0% (39/50)	96.0% (48/50)
5	94	Kernel Subspace Ensemble	Bagging_RBF_SVR_Subspace	Train MI-200	0.662 ± 0.041	[0.588, 0.729]	0.647	0.679	0.629	IN_RANGE	0.918	88.0x	88.0% (44/50)	96.0% (48/50)
6	48	Bayesian Non-Parametric	GaussianProcess_Matern_Top100	Train MI-100	0.661 ± 0.042	[0.585, 0.724]	0.647	0.656	0.651	IN_RANGE	0.907	78.0x	78.0% (39/50)	96.0% (48/50)
7	53	Histogram GBDT (Sklearn)	HistGradientBoosting_Native	Full Pruned	0.661 ± 0.033	[0.597, 0.710]	0.648	0.656	0.651	IN_RANGE	0.905	86.0x	86.0% (43/50)	100.0% (50/50)
8	99	Multi-View & Meta-Ensembles	Rank_Borda_ZScore_Hybrid_Ensemble	Full Pruned	0.660 ± 0.031	[0.595, 0.704]	0.650	0.669	0.639	IN_RANGE	0.912	94.0x	94.0% (47/50)	100.0% (50/50)
9	77	Cliff-Tolerant & DART GBDT	LightGBM_Huber_CliffTolerant	Full Pruned	0.659 ± 0.032	[0.605, 0.706]	0.650	0.657	0.650	IN_RANGE	0.907	92.0x	92.0% (46/50)	100.0% (50/50)
10	97	Multi-View & Meta-Ensembles	Simplex_NNLS_Constrained_Stacking	Full Pruned	0.659 ± 0.029	[0.613, 0.695]	0.651	0.667	0.641	IN_RANGE	0.927	94.0x	94.0% (47/50)	98.0% (49/50)
11	32	Deep Neural Network (PyTorch)	Wide_4Layer_512_256_128_64	Ext MI (<=925)	0.654 ± 0.048	[0.580, 0.715]	0.654	0.650	0.657	IN_RANGE	0.913	72.0x	72.0% (36/50)	92.0% (46/50)
12	96	Multi-View & Meta-Ensembles	TripleGBDT_XGB_LGBM_Cat_Blend	Full Pruned	0.652 ± 0.032	[0.586, 0.703]	0.657	0.655	0.652	IN_RANGE	0.906	88.0x	88.0% (44/50)	98.0% (49/50)
13	34	Deep Residual Network	Residual_256_5Blocks	Ext MI (<=925)	0.648 ± 0.040	[0.587, 0.709]	0.660	0.646	0.660	IN_RANGE	0.913	80.0x	80.0% (40/50)	94.0% (47/50)
14	82	Next-Gen Deep Tabular (PyTorch)	DualStream_PhysChem_ECFP_CrossAttn	Ext MI (<=925)	0.648 ± 0.038	[0.585, 0.697]	0.661	0.644	0.663	IN_RANGE	0.902	72.0x	72.0% (36/50)	88.0% (44/50)
15	31	Deep Neural Network (PyTorch)	Wide_3Layer_256_128_64	Ext MI (<=925)	0.644 ± 0.052	[0.558, 0.708]	0.663	0.641	0.665	IN_RANGE	0.913	68.0x	68.0% (34/50)	90.0% (45/50)
16	33	Deep Residual Network	Residual_128_3Blocks	Ext MI (<=925)	0.643 ± 0.041	[0.568, 0.707]	0.665	0.633	0.673	IN_RANGE	0.914	72.0x	72.0% (36/50)	94.0% (47/50)
17	05	Classical MLP (L1-Lasso)	MLP_Single_32_ReLU_LBFGS_L1	L1-Lasso	0.641 ± 0.031	[0.596, 0.682]	0.668	0.559	0.737	BELOW_CV	0.894	38.0x	38.0% (19/50)	84.0% (42/50)
18	60	Instance-Based (k-NN)	KNN_Manhattan_Distance_k7	Train MI-200	0.640 ± 0.053	[0.518, 0.704]	0.667	0.634	0.671	IN_RANGE	0.900	88.0x	88.0% (44/50)	100.0% (50/50)
19	45	Gradient Boosting (XGB)	XGB_Standard_Depth6	Full Pruned	0.640 ± 0.038	[0.562, 0.696]	0.668	0.649	0.657	IN_RANGE	0.904	82.0x	82.0% (41/50)	98.0% (49/50)
20	52	Oblivious GBDT (CatBoost)	CatBoost_Deep_Depth8_L2	Full Pruned	0.635 ± 0.030	[0.575, 0.687]	0.673	0.637	0.668	IN_RANGE	0.898	84.0x	84.0% (42/50)	96.0% (48/50)
21	70	Cheminformatics Similarity Kernel	KNN_Tanimoto_Jaccard_k5	Full Pruned	0.633 ± 0.049	[0.519, 0.688]	0.674	0.635	0.671	IN_RANGE	0.904	92.0x	92.0% (46/50)	100.0% (50/50)

Rank	ID	Algorithmic Family	Model Identifier	Regime	10-Fold CV R^2 (±SD)	CV [Min, Max]	CV RMSE	Val 20% R^2	Val RMSE	CV Check	ROC-AUC	EF 1% (x)	Top 1% Rec (%)	Top 5% Rec (%)
22	87	Next-Gen Deep Tabular (PyTorch)	DenseNet_Tabular_ConcatBlocks	Ext MI (<=925)	0.633 ± 0.043	[0.546, 0.688]	0.674	0.655	0.652	IN_RANGE	0.919	36.0x	36.0% (18/50)	88.0% (44/50)
23	51	Oblivious GBDT (CatBoost)	CatBoost_Oblivious_Depth6	Full Pruned	0.632 ± 0.033	[0.562, 0.680]	0.676	0.632	0.673	IN_RANGE	0.898	84.0x	84.0% (42/50)	96.0% (48/50)
24	42	Support Vector Regression	SVR_RBF_Kernel	Full Pruned	0.632 ± 0.037	[0.598, 0.706]	0.676	0.604	0.699	IN_RANGE	0.905	86.0x	86.0% (43/50)	96.0% (48/50)
25	64	Gated Highway Net	Gated_Highway_Network	Ext MI (<=925)	0.631 ± 0.024	[0.583, 0.660]	0.677	0.622	0.683	IN_RANGE	0.913	62.0x	62.0% (31/50)	92.0% (46/50)
26	38	Random Forest (Tuned)	RF_RandomizedSearch_800_Full	Full Pruned	0.628 ± 0.033	[0.560, 0.676]	0.680	0.633	0.673	IN_RANGE	0.904	92.0x	92.0% (46/50)	96.0% (48/50)
27	07	Classical MLP (L1-Lasso)	MLP_Double_32_16_ReLU_LBFGS_L1	L1-Lasso	0.627 ± 0.028	[0.581, 0.679]	0.681	0.522	0.768	BELOW_CV	0.888	26.0x	26.0% (13/50)	82.0% (41/50)
28	37	Random Forest (Bagging)	RF_Default_500_Full	Full Pruned	0.627 ± 0.032	[0.565, 0.676]	0.680	0.630	0.675	IN_RANGE	0.900	88.0x	88.0% (44/50)	98.0% (49/50)
29	83	Next-Gen Deep Tabular (PyTorch)	SwiGLU_Gated_Deep_MLP	Ext MI (<=925)	0.625 ± 0.044	[0.534, 0.691]	0.682	0.665	0.642	IN_RANGE	0.907	66.0x	66.0% (33/50)	94.0% (47/50)
30	71	Instance-Based (k-NN)	KNN_Cosine_Similarity_k9	Train MI-200	0.624 ± 0.044	[0.552, 0.683]	0.682	0.634	0.672	IN_RANGE	0.907	42.0x	42.0% (21/50)	72.0% (36/50)
31	46	Gradient Boosting (XGB)	XGB_Deep_Depth12	Full Pruned	0.624 ± 0.032	[0.559, 0.678]	0.683	0.623	0.681	IN_RANGE	0.892	84.0x	84.0% (42/50)	100.0% (50/50)
32	73	Cliff-Tolerant & DART GBDT	XGB_DART_DropoutTrees	Full Pruned	0.619 ± 0.032	[0.549, 0.662]	0.688	0.620	0.684	IN_RANGE	0.899	82.0x	82.0% (41/50)	98.0% (49/50)
33	80	Cliff-Tolerant & DART GBDT	HistGBDT_L1_AbsoluteError	Full Pruned	0.614 ± 0.027	[0.560, 0.659]	0.692	0.625	0.680	IN_RANGE	0.889	84.0x	84.0% (42/50)	98.0% (49/50)
34	54	Adaptive Boosting	AdaBoost_ExtraTrees_Ensemble	Full Pruned	0.613 ± 0.034	[0.552, 0.659]	0.693	0.624	0.680	IN_RANGE	0.906	88.0x	88.0% (44/50)	94.0% (47/50)
35	75	Cliff-Tolerant & DART GBDT	XGB_L1_AbsoluteError	Full Pruned	0.612 ± 0.034	[0.549, 0.657]	0.694	0.617	0.687	IN_RANGE	0.895	78.0x	78.0% (39/50)	96.0% (48/50)
36	92	Sparse & Robust Linear/Latent	RidgeCV_MultiAlpha_Full	Full Pruned	0.610 ± 0.030	[0.583, 0.680]	0.696	0.611	0.693	IN_RANGE	0.905	56.0x	56.0% (28/50)	86.0% (43/50)
37	30	Deep Neural Network (PyTorch)	Double_BN_ELU_96_48	Ext MI (<=925)	0.609 ± 0.045	[0.535, 0.669]	0.696	0.630	0.675	IN_RANGE	0.912	72.0x	72.0% (36/50)	96.0% (48/50)
38	78	Cliff-Tolerant & DART GBDT	CatBoost_MAE_CliffTolerant	Full Pruned	0.608 ± 0.029	[0.546, 0.653]	0.698	0.622	0.682	IN_RANGE	0.894	76.0x	76.0% (38/50)	96.0% (48/50)
39	81	Next-Gen Deep Tabular (PyTorch)	FT_Transformer_Token_Attention	Ext MI (<=925)	0.608 ± 0.043	[0.541, 0.669]	0.697	0.575	0.724	IN_RANGE	0.902	60.0x	60.0% (30/50)	90.0% (45/50)
40	06	Classical MLP (L1-Lasso)	MLP_Single_32_Tanh_LBFGS_L1	L1-Lasso	0.605 ± 0.030	[0.564, 0.666]	0.700	0.540	0.753	BELOW_CV	0.898	54.0x	54.0% (27/50)	90.0% (45/50)

Rank	ID	Algorithmic Family	Model Identifier	Regime	10-Fold CV R^2 (±SD)	CV [Min, Max]	CV RMSE	Val 20% R^2	Val RMSE	CV Check	ROC-AUC	EF 1% (x)	Top 1% Rec (%)	Top 5% Rec (%)
41	36	Deep Bayesian / Tuned MLP	Optuna_Tuned_BN_MLP	Ext MI (<=925)	0.598 ± 0.045	[0.527, 0.682]	0.705	0.625	0.680	IN_RANGE	0.912	76.0x	76.0% (38/50)	96.0% (48/50)
42	95	Multi-View & Meta-Ensembles	LateFusion_RF_plus_TanimotoKRR	Full Pruned	0.592 ± 0.028	[0.533, 0.633]	0.712	0.603	0.700	IN_RANGE	0.898	94.0x	94.0% (47/50)	100.0% (50/50)
43	66	Cheminformatics Similarity Kernel	SVR_Tanimoto_Similarity_Kernel	Full Pruned	0.587 ± 0.037	[0.518, 0.638]	0.716	0.592	0.709	IN_RANGE	0.890	80.0x	80.0% (40/50)	96.0% (48/50)
44	72	Instance-Based (k-NN)	KNN_Euclidean_L2_k11	Train MI-200	0.582 ± 0.045	[0.490, 0.653]	0.719	0.596	0.706	IN_RANGE	0.893	56.0x	56.0% (28/50)	90.0% (45/50)
45	69	Cheminformatics Similarity Kernel	GPR_Tanimoto_Covariance_Kernel	Train MI-200	0.568 ± 0.039	[0.473, 0.614]	0.732	0.573	0.726	IN_RANGE	0.876	72.0x	72.0% (36/50)	94.0% (47/50)
46	08	Classical MLP (L1-Lasso)	MLP_Double_32_16_Tanh_LBFGS_L1	L1-Lasso	0.542 ± 0.046	[0.437, 0.581]	0.754	0.471	0.808	IN_RANGE	0.889	24.0x	24.0% (12/50)	72.0% (36/50)
47	98	Multi-View & Meta-Ensembles	Huber_CliffRobust_Meta_Stacking	Full Pruned	0.533 ± 0.031	[0.478, 0.588]	0.762	0.557	0.739	IN_RANGE	0.892	52.0x	52.0% (26/50)	92.0% (46/50)
48	43	Support Vector Regression	SVR_Sigmoid_Kernel	Full Pruned	0.529 ± 0.034	[0.492, 0.618]	0.764	0.474	0.805	IN_RANGE	0.874	4.0x	4.0% (2/50)	58.0% (29/50)
49	65	7-Family Ultra-Stacking	Ultra_Stacking_7Family_RidgeCV	Full Pruned	0.501 ± 0.034	[0.430, 0.560]	0.788	0.523	0.767	IN_RANGE	0.882	68.0x	68.0% (34/50)	96.0% (48/50)
50	50	Champion Super-Stacking	Super_Stacking_RidgeCV	Full Pruned	0.499 ± 0.033	[0.434, 0.553]	0.789	0.521	0.768	IN_RANGE	0.881	76.0x	76.0% (38/50)	96.0% (48/50)
51	62	Deep 1D ConvNet	CNN_1D_Pharmacophore_Motif	Ext MI (<=925)	0.496 ± 0.040	[0.411, 0.563]	0.790	0.491	0.792	IN_RANGE	0.844	46.0x	46.0% (23/50)	84.0% (42/50)
52	22	Classical MLP (MI-150)	MLP_Single_32_Tanh_LBFGS_MI150	Train MI-150	0.494 ± 0.068	[0.343, 0.575]	0.790	0.489	0.793	IN_RANGE	0.854	54.0x	54.0% (27/50)	90.0% (45/50)
53	68	Cheminformatics Similarity Kernel	KRR_MinMax_Pharmacophore_Kernel	Full Pruned	0.491 ± 0.025	[0.444, 0.520]	0.796	0.504	0.782	IN_RANGE	0.863	82.0x	82.0% (41/50)	98.0% (49/50)
54	21	Classical MLP (MI-150)	MLP_Single_32_ReLU_LBFGS_MI150	Train MI-150	0.487 ± 0.050	[0.417, 0.580]	0.797	0.499	0.786	IN_RANGE	0.874	38.0x	38.0% (19/50)	84.0% (42/50)
55	29	Deep Neural Network (PyTorch)	Single_BN_Swish_128	Ext MI (<=925)	0.476 ± 0.061	[0.416, 0.620]	0.806	0.482	0.799	IN_RANGE	0.895	86.0x	86.0% (43/50)	94.0% (47/50)
56	57	Bayesian Linear ARD	BayesianRidge_EmpiricalBayes	Train MI-200	0.475 ± 0.048	[0.375, 0.555]	0.807	0.465	0.812	IN_RANGE	0.842	48.0x	48.0% (24/50)	90.0% (45/50)
57	67	Cheminformatics Similarity Kernel	KRR_Tanimoto_Similarity_Kernel	Full Pruned	0.475 ± 0.026	[0.428, 0.506]	0.808	0.491	0.792	IN_RANGE	0.859	82.0x	82.0% (41/50)	98.0% (49/50)
58	41	Support Vector Regression	SVR_Poly_Degree3	Full Pruned	0.472 ± 0.041	[0.390, 0.522]	0.810	0.493	0.790	IN_RANGE	0.878	66.0x	66.0% (33/50)	80.0% (40/50)
59	44	Regularized Linear / KRR	ElasticNetCV_Top200	Train MI-200	0.470 ± 0.047	[0.375, 0.542]	0.811	0.463	0.814	IN_RANGE	0.843	46.0x	46.0% (23/50)	88.0% (44/50)
60	59	Latent Variable QSAR	PLS_Regression_10Latent	Train MI-200	0.466 ± 0.050	[0.361, 0.546]	0.814	0.447	0.825	IN_RANGE	0.833	46.0x	46.0% (23/50)	86.0% (43/50)

Rank	ID	Algorithmic Family	Model Identifier	Regime	10-Fold CV R^2 (±SD)	CV [Min, Max]	CV RMSE	Val 20% R^2	Val RMSE	CV Check	ROC-AUC	EF 1% (x)	Top 1% Rec (%)	Top 5% Rec (%)
61	86	Next-Gen Deep Tabular (PyTorch)	Cosine_WeightNorm_Swish_MLP	Ext MI (<=925)	0.465 ± 0.071	[0.303, 0.581]	0.813	0.412	0.851	IN_RANGE	0.907	74.0x	74.0% (37/50)	94.0% (47/50)
62	58	Robust Linear Regression	HuberRegressor_OutlierRobust	Train MI-200	0.453 ± 0.049	[0.360, 0.532]	0.823	0.454	0.820	IN_RANGE	0.841	44.0x	44.0% (22/50)	86.0% (43/50)
63	24	Classical MLP (MI-150)	MLP_Double_32_16_Tanh_LBFGS_MI150	Train MI-150	0.444 ± 0.109	[0.277, 0.583]	0.826	0.445	0.827	IN_RANGE	0.856	24.0x	24.0% (12/50)	72.0% (36/50)
64	19	Classical MLP (MI-150)	MLP_Double_32_16_ReLU_Adam_MI150	Train MI-150	0.441 ± 0.077	[0.287, 0.537]	0.831	0.541	0.752	IN_RANGE	0.884	4.0x	4.0% (2/50)	64.0% (32/50)
65	14	Classical MLP (PCA-150)	MLP_Single_32_Tanh_LBFGS_PCA	PCA-150	0.421 ± 0.065	[0.285, 0.489]	0.847	0.487	0.795	IN_RANGE	0.860	42.0x	42.0% (21/50)	76.0% (38/50)
66	100	10-Family Grand Ultra-Stacking	Grand_10Family_Ultra_Stacking	Full Pruned	0.414 ± 0.055	[0.311, 0.489]	0.853	0.442	0.829	IN_RANGE	0.847	40.0x	40.0% (20/50)	58.0% (29/50)
67	89	Sparse & Robust Linear/Latent	OMP_Sparse_Pharmacophore_50	Train MI-200	0.411 ± 0.055	[0.285, 0.505]	0.855	0.398	0.861	IN_RANGE	0.812	46.0x	46.0% (23/50)	88.0% (44/50)
68	23	Classical MLP (MI-150)	MLP_Double_32_16_ReLU_LBFGS_MI150	Train MI-150	0.410 ± 0.097	[0.239, 0.602]	0.853	0.453	0.821	IN_RANGE	0.860	26.0x	26.0% (13/50)	82.0% (41/50)
69	93	Sparse & Robust Linear/Latent	PLS_Regression_25Latent_Full	Full Pruned	0.386 ± 0.058	[0.309, 0.499]	0.873	0.470	0.808	IN_RANGE	0.882	2.0x	2.0% (1/50)	56.0% (28/50)
70	35	Self-Attention Network	Attention_128_4Heads	Ext MI (<=925)	0.377 ± 0.098	[0.191, 0.519]	0.878	0.496	0.788	IN_RANGE	0.885	62.0x	62.0% (31/50)	84.0% (42/50)
71	76	Cliff-Tolerant & DART GBDT	LightGBM_DART_DropoutMode	Full Pruned	0.376 ± 0.054	[0.281, 0.463]	0.879	0.373	0.879	IN_RANGE	0.894	84.0x	84.0% (42/50)	96.0% (48/50)
72	17	Classical MLP (MI-150)	MLP_Single_32_ReLU_Adam_MI150	Train MI-150	0.358 ± 0.112	[0.131, 0.543]	0.888	0.445	0.827	IN_RANGE	0.861	36.0x	36.0% (18/50)	84.0% (42/50)
73	15	Classical MLP (PCA-150)	MLP_Double_32_16_ReLU_LBFGS_PCA	PCA-150	0.324 ± 0.116	[0.089, 0.468]	0.912	0.456	0.818	IN_RANGE	0.882	22.0x	22.0% (11/50)	74.0% (37/50)
74	16	Classical MLP (PCA-150)	MLP_Double_32_16_Tanh_LBFGS_PCA	PCA-150	0.279 ± 0.094	[0.100, 0.388]	0.944	0.320	0.915	IN_RANGE	0.836	12.0x	12.0% (6/50)	58.0% (29/50)
75	18	Classical MLP (MI-150)	MLP_Single_32_Tanh_Adam_MI150	Train MI-150	0.271 ± 0.087	[0.072, 0.354]	0.950	0.448	0.825	ABOVE_CV	0.869	54.0x	54.0% (27/50)	82.0% (41/50)
76	56	Kernel Ridge (Laplacian)	KRR_Laplacian_CityBlock	Full Pruned	0.267 ± 0.113	[0.045, 0.511]	0.951	0.241	0.967	IN_RANGE	0.828	70.0x	70.0% (35/50)	96.0% (48/50)
77	28	Deep Neural Network (PyTorch)	Double_Regression_64_32_PyTorch	Ext MI (<=925)	0.231 ± 0.115	[0.032, 0.368]	0.975	0.344	0.899	IN_RANGE	0.878	64.0x	64.0% (32/50)	88.0% (44/50)
78	85	Next-Gen Deep Tabular (PyTorch)	Heteroscedastic_GaussianNLL_Net	Ext MI (<=925)	0.209 ± 0.090	[0.075, 0.353]	0.989	0.312	0.921	IN_RANGE	0.910	74.0x	74.0% (37/50)	96.0% (48/50)
79	13	Classical MLP (PCA-150)	MLP_Single_32_ReLU_LBFGS_PCA	PCA-150	0.209 ± 0.212	[-0.155, 0.478]	0.981	0.311	0.922	IN_RANGE	0.869	30.0x	30.0% (15/50)	72.0% (36/50)
80	91	Sparse & Robust Linear/Latent	QuantileRegressor_L1_Median	Train MI-200	0.208 ± 0.020	[0.182, 0.241]	0.992	0.210	0.987	IN_RANGE	0.733	32.0x	32.0% (16/50)	62.0% (31/50)

Rank	ID	Algorithmic Family	Model Identifier	Regime	10-Fold CV R^2 (±SD)	CV [Min, Max]	CV RMSE	Val 20% R^2	Val RMSE	CV Check	ROC-AUC	EF 1% (x)	Top 1% Rec (%)	Top 5% Rec (%)
81	20	Classical MLP (MI-150)	MLP_Double_32_16_Tanh_Adam_MI150	Train MI-150	0.208 ± 0.130	[-0.026, 0.414]	0.987	0.511	0.776	ABOVE_CV	0.889	44.0x	44.0% (22/50)	82.0% (41/50)
82	90	Sparse & Robust Linear/Latent	LarsCV_LeastAngle_Path	Train MI-100	0.201 ± 0.048	[0.083, 0.280]	0.997	0.223	0.978	IN_RANGE	0.729	42.0x	42.0% (21/50)	88.0% (44/50)
83	88	Next-Gen Deep Tabular (PyTorch)	MultiScale_Dilated_1DCNN	Ext MI (<=925)	0.180 ± 0.038	[0.129, 0.237]	1.009	0.228	0.975	IN_RANGE	0.715	0.0x	0.0% (0/50)	2.0% (1/50)
84	27	Deep Neural Network (PyTorch)	Double_Regression_32_16_PyTorch	Ext MI (<=925)	0.092 ± 0.165	[-0.189, 0.290]	1.058	0.257	0.957	IN_RANGE	0.878	28.0x	28.0% (14/50)	84.0% (42/50)
85	11	Classical MLP (PCA-150)	MLP_Double_32_16_ReLU_Adam_PCA	PCA-150	-0.058 ± 0.208	[-0.543, 0.194]	1.140	0.313	0.920	IN_RANGE	0.855	40.0x	40.0% (20/50)	90.0% (45/50)
86	26	Deep Neural Network (PyTorch)	Single_Regression_64_PyTorch	Ext MI (<=925)	-0.171 ± 0.222	[-0.753, 0.060]	1.201	0.283	0.940	ABOVE_CV	0.877	78.0x	78.0% (39/50)	92.0% (46/50)
87	09	Classical MLP (PCA-150)	MLP_Single_32_ReLU_Adam_PCA	PCA-150	-0.227 ± 0.236	[-0.847, 0.095]	1.232	-0.005	1.113	IN_RANGE	0.837	28.0x	28.0% (14/50)	82.0% (41/50)
88	25	Deep Neural Network (PyTorch)	Single_Regression_32_PyTorch	Ext MI (<=925)	-0.384 ± 0.249	[-0.954, -0.023]	1.308	0.090	1.059	IN_RANGE	0.855	56.0x	56.0% (28/50)	90.0% (45/50)
89	12	Classical MLP (PCA-150)	MLP_Double_32_16_Tanh_Adam_PCA	PCA-150	-0.754 ± 0.264	[-1.167, -0.285]	1.471	-0.062	1.144	ABOVE_CV	0.788	40.0x	40.0% (20/50)	74.0% (37/50)
90	10	Classical MLP (PCA-150)	MLP_Single_32_Tanh_Adam_PCA	PCA-150	-0.812 ± 0.387	[-1.616, -0.349]	1.493	-0.314	1.273	IN_RANGE	0.795	26.0x	26.0% (13/50)	76.0% (38/50)
91	03	Classical MLP (L1-Lasso)	MLP_Double_32_16_ReLU_Adam_L1	L1-Lasso	-0.842 ± 0.239	[-1.313, -0.440]	1.509	-0.126	1.178	ABOVE_CV	0.823	4.0x	4.0% (2/50)	64.0% (32/50)
92	63	Self-Normalizing NN	SNN_SELU_AlphaDropout	Ext MI (<=925)	-1.890 ± 0.909	[-4.400, -1.038]	1.878	-0.872	1.519	IN_RANGE	0.873	70.0x	70.0% (35/50)	100.0% (50/50)
93	40	Support Vector Regression	LinearSVR_Fast	Full Pruned	-2.125 ± 0.527	[-3.249, -1.152]	1.964	0.098	1.054	ABOVE_CV	0.859	6.0x	6.0% (3/50)	54.0% (27/50)
94	55	Kernel Ridge Regression	KRR_RBF_Kernel	Full Pruned	-3.400 ± 0.346	[-3.980, -2.549]	2.336	-3.685	2.403	IN_RANGE	0.711	36.0x	36.0% (18/50)	78.0% (39/50)
95	01	Classical MLP (L1-Lasso)	MLP_Single_32_ReLU_Adam_L1	L1-Lasso	-3.659 ± 0.737	[-4.898, -2.786]	2.404	-0.369	1.299	ABOVE_CV	0.813	36.0x	36.0% (18/50)	84.0% (42/50)
96	79	Cliff-Tolerant & DART GBDT	CatBoost_Huber_Loss	Full Pruned	-4.358 ± 3.726	[-12.651, 0.033]	2.410	-0.589	1.399	IN_RANGE	0.810	2.0x	2.0% (1/50)	68.0% (34/50)
97	04	Classical MLP (L1-Lasso)	MLP_Double_32_16_Tanh_Adam_L1	L1-Lasso	-9.413 ± 1.160	[-11.393, -7.143]	3.594	-2.207	1.988	ABOVE_CV	0.778	44.0x	44.0% (22/50)	82.0% (41/50)
98	02	Classical MLP (L1-Lasso)	MLP_Single_32_Tanh_Adam_L1	L1-Lasso	-16.973 ± 2.441	[-22.571, -13.171]	4.721	-1.746	1.839	ABOVE_CV	0.782	54.0x	54.0% (27/50)	82.0% (41/50)
99	84	Next-Gen Deep Tabular (PyTorch)	SparseMask_TabNet_SequentialAttn	Ext MI (<=925)	-126.205 ± 252.097	[-800.884, 0.583]	7.153	-23.888	5.538	IN_RANGE	0.903	70.0x	70.0% (35/50)	90.0% (45/50)

Rank	ID	Algorithmic Family	Model Identifier	Regime	10-Fold CV R^2 (±SD)	CV [Min, Max]	CV RMSE	Val 20% R^2	Val RMSE	CV Check	ROC-AUC	EF 1% (x)	Top 1% Rec (%)	Top 5% Rec (%)
100	74	Cliff-Tolerant & DART GBDT	XGB_PseudoHuber_CliffRobust	Full Pruned	-66975.172 ± 4655.355	[-76504.280, -57784.548]	288.090	-67354.313	288.088	IN_RANGE	0.500	4.0x	4.0% (2/50)	60.0% (30/50)

S3. Rearranged High-Reliability Models (10-Fold CV R^2 > 0.50) Ranked by Top 1% Decoy Recovery

Supplementary Table S3 illustrates the two-stage selection rearrangement (Sheet 2 protocol).

Supplementary Table S3. Rearranged subset of models from the 100-model suite satisfying 10-Fold CV R^2 > 0.50 ordered by Top 1% Decoy Active Recovery.

Selection Rank	ID	Model Identifier	Algorithmic Family	Regime	Top 1% Recovery (%)	EF 1% (x)	10-Fold CV R^2 (±SD)	Val 20% R^2	CV Check	Top 5% Recovery (%)
Top-1*	39	ExtraTrees_500_Full	Extremely Randomized Trees	Full Pruned	94.0% (47/50)	94.0x	0.686 ± 0.034	0.704	IN_RANGE	98.0%
Top-2*	49	Ensemble_Voting_RF_ET_XGB	Heterogeneous Ensembles	Full Pruned	94.0% (47/50)	94.0x	0.668 ± 0.033	0.680	IN_RANGE	100.0%
Top-3*	99	Rank_Borda_ZScore_Hybrid_Ensemble	Multi-View & Meta-Ensembles	Full Pruned	94.0% (47/50)	94.0x	0.660 ± 0.031	0.669	IN_RANGE	100.0%
Top-4*	97	Simplex_NNLS_Constrained_Stacking	Multi-View & Meta-Ensembles	Full Pruned	94.0% (47/50)	94.0x	0.659 ± 0.029	0.667	IN_RANGE	98.0%
Top-5*	95	LateFusion_RF_plus_TanimotoKRR	Multi-View & Meta-Ensembles	Full Pruned	94.0% (47/50)	94.0x	0.592 ± 0.028	0.603	IN_RANGE	100.0%
6	77	LightGBM_Huber_CliffTolerant	Cliff-Tolerant & DART GBDT	Full Pruned	92.0% (46/50)	92.0x	0.659 ± 0.032	0.657	IN_RANGE	100.0%
7	70	KNN_Tanimoto_Jaccard_k5	Cheminformatics Similarity Kernel	Full Pruned	92.0% (46/50)	92.0x	0.633 ± 0.049	0.635	IN_RANGE	100.0%
8	38	RF_RandomizedSearch_800_Full	Random Forest (Tuned)	Full Pruned	92.0% (46/50)	92.0x	0.628 ± 0.033	0.633	IN_RANGE	96.0%
9	47	LightGBM_Hist_Full	Histogram / Oblivious GBDT	Full Pruned	90.0% (45/50)	90.0x	0.665 ± 0.031	0.660	IN_RANGE	100.0%
10	94	Bagging_RBF_SVR_Subspace	Kernel Subspace Ensemble	Train MI-200	88.0% (44/50)	88.0x	0.662 ± 0.041	0.679	IN_RANGE	96.0%
11	96	TripleGBDT_XGB_LGBM_Cat_Blend	Multi-View & Meta-Ensembles	Full Pruned	88.0% (44/50)	88.0x	0.652 ± 0.032	0.655	IN_RANGE	98.0%
12	60	KNN_Manhattan_Distance_k7	Instance-Based (k-NN)	Train MI-200	88.0% (44/50)	88.0x	0.640 ± 0.053	0.634	IN_RANGE	100.0%

Selection Rank	ID	Model Identifier	Algorithmic Family	Regime	Top 1% Recovery (%)	EF 1% (x)	10-Fold CV R^2 (±SD)	Val 20% R^2	CV Check	Top 5% Recovery (%)
13	37	RF_Default_500_Full	Random Forest (Bagging)	Full Pruned	88.0% (44/50)	88.0x	0.627 ± 0.032	0.630	IN_RANGE	98.0%
14	54	AdaBoost_ExtraTrees_Ensemble	Adaptive Boosting	Full Pruned	88.0% (44/50)	88.0x	0.613 ± 0.034	0.624	IN_RANGE	94.0%
15	53	HistGradientBoosting_Native	Histogram GBDT (Sklearn)	Full Pruned	86.0% (43/50)	86.0x	0.661 ± 0.033	0.656	IN_RANGE	100.0%
16	42	SVR_RBF_Kernel	Support Vector Regression	Full Pruned	86.0% (43/50)	86.0x	0.632 ± 0.037	0.604	IN_RANGE	96.0%
17	52	CatBoost_Deep_Depth8_L2	Oblivious GBDT (CatBoost)	Full Pruned	84.0% (42/50)	84.0x	0.635 ± 0.030	0.637	IN_RANGE	96.0%
18	51	CatBoost_Oblivious_Depth6	Oblivious GBDT (CatBoost)	Full Pruned	84.0% (42/50)	84.0x	0.632 ± 0.033	0.632	IN_RANGE	96.0%
19	46	XGB_Deep_Depth12	Gradient Boosting (XGB)	Full Pruned	84.0% (42/50)	84.0x	0.624 ± 0.032	0.623	IN_RANGE	100.0%
20	80	HistGBDT_L1_AbsoluteError	Cliff-Tolerant & DART GBDT	Full Pruned	84.0% (42/50)	84.0x	0.614 ± 0.027	0.625	IN_RANGE	98.0%
21	45	XGB_Standard_Depth6	Gradient Boosting (XGB)	Full Pruned	82.0% (41/50)	82.0x	0.640 ± 0.038	0.649	IN_RANGE	98.0%
22	73	XGB_DART_DropoutTrees	Cliff-Tolerant & DART GBDT	Full Pruned	82.0% (41/50)	82.0x	0.619 ± 0.032	0.620	IN_RANGE	98.0%
23	34	Residual_256_5Blocks	Deep Residual Network	Ext MI (<=925)	80.0% (40/50)	80.0x	0.648 ± 0.040	0.646	IN_RANGE	94.0%
24	66	SVR_Tanimoto_Similarity_Kernel	Cheminformatics Similarity Kernel	Full Pruned	80.0% (40/50)	80.0x	0.587 ± 0.037	0.592	IN_RANGE	96.0%
25	61	GP_RationalQuadratic_Top100	Bayesian Multi-Scale GP	Train MI-100	78.0% (39/50)	78.0x	0.663 ± 0.042	0.658	IN_RANGE	96.0%
26	48	GaussianProcess_Matern_Top100	Bayesian Non-Parametric	Train MI-100	78.0% (39/50)	78.0x	0.661 ± 0.042	0.656	IN_RANGE	96.0%
27	75	XGB_L1_AbsoluteError	Cliff-Tolerant & DART GBDT	Full Pruned	78.0% (39/50)	78.0x	0.612 ± 0.034	0.617	IN_RANGE	96.0%
28	78	CatBoost_MAE_CliffTolerant	Cliff-Tolerant & DART GBDT	Full Pruned	76.0% (38/50)	76.0x	0.608 ± 0.029	0.622	IN_RANGE	96.0%
29	36	Optuna_Tuned_BN_MLP	Deep Bayesian / Tuned MLP	Ext MI (<=925)	76.0% (38/50)	76.0x	0.598 ± 0.045	0.625	IN_RANGE	96.0%
30	32	Wide_4Layer_512_256_128_64	Deep Neural Network (PyTorch)	Ext MI (<=925)	72.0% (36/50)	72.0x	0.654 ± 0.048	0.650	IN_RANGE	92.0%
31	82	DualStream_PhysChem_ECFP_CrossAttn	Next-Gen Deep Tabular (PyTorch)	Ext MI (<=925)	72.0% (36/50)	72.0x	0.648 ± 0.038	0.644	IN_RANGE	88.0%

Selection Rank	ID	Model Identifier	Algorithmic Family	Regime	Top 1% Recovery (%)	EF 1% (x)	10-Fold CV R^2 (±SD)	Val 20% R^2	CV Check	Top 5% Recovery (%)
32	33	Residual_128_3Blocks	Deep Residual Network	Ext MI (<=925)	72.0% (36/50)	72.0x	0.643 ± 0.041	0.633	IN_RANGE	94.0%
33	30	Double_BN_ELU_96_48	Deep Neural Network (PyTorch)	Ext MI (<=925)	72.0% (36/50)	72.0x	0.609 ± 0.045	0.630	IN_RANGE	96.0%
34	69	GPR_Tanimoto_Covariance_Kernel	Cheminformatics Similarity Kernel	Train MI-200	72.0% (36/50)	72.0x	0.568 ± 0.039	0.573	IN_RANGE	94.0%
35	31	Wide_3Layer_256_128_64	Deep Neural Network (PyTorch)	Ext MI (<=925)	68.0% (34/50)	68.0x	0.644 ± 0.052	0.641	IN_RANGE	90.0%
36	65	Ultra_Stacking_7Family_RidgeCV	7-Family Ultra-Stacking	Full Pruned	68.0% (34/50)	68.0x	0.501 ± 0.034	0.523	IN_RANGE	96.0%
37	83	SwiGLU_Gated_Deep_MLP	Next-Gen Deep Tabular (PyTorch)	Ext MI (<=925)	66.0% (33/50)	66.0x	0.625 ± 0.044	0.665	IN_RANGE	94.0%
38	64	Gated_Highway_Network	Gated Highway Net	Ext MI (<=925)	62.0% (31/50)	62.0x	0.631 ± 0.024	0.622	IN_RANGE	92.0%
39	81	FT_Transformer_Token_Attention	Next-Gen Deep Tabular (PyTorch)	Ext MI (<=925)	60.0% (30/50)	60.0x	0.608 ± 0.043	0.575	IN_RANGE	90.0%
40	92	RidgeCV_MultiAlpha_Full	Sparse & Robust Linear/Latent	Full Pruned	56.0% (28/50)	56.0x	0.610 ± 0.030	0.611	IN_RANGE	86.0%
41	72	KNN_Euclidean_L2_k11	Instance-Based (k-NN)	Train MI-200	56.0% (28/50)	56.0x	0.582 ± 0.045	0.596	IN_RANGE	90.0%
42	06	MLP_Single_32_Tanh_LBFGS_L1	Classical MLP (L1-Lasso)	L1-Lasso	54.0% (27/50)	54.0x	0.605 ± 0.030	0.540	BELOW_CV	90.0%
43	98	Huber_CliffRobust_Meta_Stacking	Multi-View & Meta-Ensembles	Full Pruned	52.0% (26/50)	52.0x	0.533 ± 0.031	0.557	IN_RANGE	92.0%
44	71	KNN_Cosine_Similarity_k9	Instance-Based (k-NN)	Train MI-200	42.0% (21/50)	42.0x	0.624 ± 0.044	0.634	IN_RANGE	72.0%
45	05	MLP_Single_32_ReLU_LBFGS_L1	Classical MLP (L1-Lasso)	L1-Lasso	38.0% (19/50)	38.0x	0.641 ± 0.031	0.559	BELOW_CV	84.0%
46	87	DenseNet_Tabular_ConcatBlocks	Next-Gen Deep Tabular (PyTorch)	Ext MI (<=925)	36.0% (18/50)	36.0x	0.633 ± 0.043	0.655	IN_RANGE	88.0%
47	07	MLP_Double_32_16_ReLU_LBFGS_L1	Classical MLP (L1-Lasso)	L1-Lasso	26.0% (13/50)	26.0x	0.627 ± 0.028	0.522	BELOW_CV	82.0%
48	08	MLP_Double_32_16_Tanh_LBFGS_L1	Classical MLP (L1-Lasso)	L1-Lasso	24.0% (12/50)	24.0x	0.542 ± 0.046	0.471	IN_RANGE	72.0%
49	43	SVR_Sigmoid_Kernel	Support Vector Regression	Full Pruned	4.0% (2/50)	4.0x	0.529 ± 0.034	0.474	IN_RANGE	58.0%

S4. Abbreviations and Technical Nomenclature

Supplementary Table S4. Abbreviations, technical nomenclature, and official target synonyms used throughout the manuscript.

Abbreviation / Symbol	Full Scientific Term / Mathematical Definition & Target Synonyms
ADME	Absorption, Distribution, Metabolism, and Excretion
ARD	Automatic Relevance Determination
ATP	Adenosine Triphosphate
BM	Bemis-Murcko Molecular Scaffold Framework
BN	Batch Normalization
CAIR	Chemical Artificial Intelligence Reporting Guidelines
ChEMBL	Chemical European Molecular Biology Database
CNN	Convolutional Neural Network
CV	Cross-Validation (10-Fold Stratified)
DART	Dropouts Meet Multiple Additive Regression Trees
DFG	Asp-Phe-Gly Conserved Kinase Activation Loop Motif
ECFP	Extended-Connectivity Fingerprints (Morgan Algorithm)
EF	Enrichment Factor (Virtual Screening Decoy Recovery Metric)
ELU	Exponential Linear Unit
ET	Extremely Randomized Trees (ExtraTrees)
FT-Transformer	Feature Tokenizer Transformer for Tabular Data
GBDT	Gradient Boosted Decision Trees
GP / GPR	Gaussian Process Regression
IC50	Half-Maximal Inhibitory Concentration
k-NN	k-Nearest Neighbors Regression
KLIFS	Kinase-Ligand Interaction Fingerprints and Structures Database
KRR	Kernel Ridge Regression
L-BFGS	Limited-Memory Broyden-Fletcher-Goldfarb-Shanno Optimizer
LARS	Least Angle Regression
LGBM	Light Gradient Boosting Machine
MAE	Mean Absolute Error
MAPK14 (CHEMBL260)	Mitogen-activated protein kinase 14; Official Synonyms: CSBP, CSBP1, CSBP2, CSPB1, EXIP, MAP kinase 14, MAP kinase MXI2, MAPK 14, MAPK14, Mxi2, p38, p38 MAP kinase, p38a, p38α, p38α Exip, PRKM14, PRKM15, RK
MCC	Matthews Correlation Coefficient
MI	Mutual Information (Shannon Entropy-Based Dependency)
MLP	Multilayer Perceptron
NND	Nearest-Neighbor Tanimoto Distance
NNLS	Non-Negative Least Squares
OMP	Orthogonal Matching Pursuit
PCA	Principal Component Analysis
pIC50	Negative Base-10 Logarithm of Molar IC50 (-log10[IC50 in M])
PLS	Partial Least Squares Regression
QSAR	Quantitative Structure-Activity Relationship

Abbreviation / Symbol	Full Scientific Term / Mathematical Definition & Target Synonyms
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
ResNet	Deep Residual Neural Network
RF	Random Forest
RMSE	Root Mean Squared Error
ROC-AUC	Area Under the Receiver Operating Characteristic Curve
R^2	Coefficient of Determination
SAR	Structure-Activity Relationship
SD	Standard Deviation
SELU	Scaled Exponential Linear Unit
SiLU	Sigmoid Linear Unit (Swish Activation)
SMILES	Simplified Molecular-Input Line-Entry System
SNN	Self-Normalizing Neural Network
SVR	Support Vector Regression
SwiGLU	Swish-Gated Linear Unit
XGB / XGBoost	Extreme Gradient Boosting

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MAPK14SUPPLEMENTARY.zip