Predicting Super-Class Drug Mechanisms from Large-Scale IC50 Cell-Line Sensitivity Profiles

Tauhidur Rahman Sakib[*1], Mostofa Adib Shakib[2], Saad Ibn Munir[3], Saiful Islam Badhon[4], Md. Sadman Yeasir Sakif[1] and Mim Deb Roy[1]

[1]*BRAC University, Dhaka-1212, Bangladesh*
[2]*Texas Tech University, United States (USA)*
[3]*University of Malaya, Kuala Lumpur, Malaysia*
[4]*California State University, East Bay, United States (USA)*

Email- tauhidurrahmansakib@gmail.com, adibshakib@gmail.com, saadibnmunir@gmail.com, mdsaifulbadhon2002@gmail.com, sadmanyeasir123@gmail.com, mim.deb.roy@g.bracu.ac.bd

***Abstract*—Predicting mechanisms of action of drug compounds at the super-class level is indispensable for early drug discovery and functional assessment of compounds. The aim of this study is to perform Super-Class–level mechanism-of-action prediction from phenotypic IC50 cell-line sensitivity data under strict leakage control. In this work, we present a leakage-safe machine learning pipeline for Super-Class Mechanism-of-Action (Super-Class MoA) prediction using drug–cell line sensitivity profiles prepared at large scale from the GDSC dataset. The standardized features derived from IC50 were encoded into 32-dimensional latent embeddings using an autoencoder, while SMOTE was applied solely to the training embeddings of each outer fold to address class imbalance. The resulting embeddings were used to train a stacked neural-boosted ensemble consisting of a neural network and XGBoost, with validation embeddings kept completely untouched to ensure unbiased evaluation. The framework achieved a mean accuracy of 0.9182, a mean macro-F1 score of 0.9118, and a mean macro-AUROC of 0.9568 across five outer folds, demonstrating consistent performance across the three Super-Class MoA categories. These results indicate that latent representations derived from phenotypic drug-response profiles capture biologically meaningful variation, supporting robust Super-Class MoA prediction in early-stage screening settings.**

Keywords— MoA, Latent features, Autoencoder, Embeddings, IC50, Resampling, Phenotypic data, SMOTE, Ensemble.

I. Introduction

Predicting Predicting Super-Class mechanisms of action (MoA) directly from IC50 profiles is appealing because these assays are inexpensive and widely available; however, the task is challenged by noise, nonlinear response patterns, and strong class imbalance. Most existing approaches rely on chemical structures or transcriptomic data, leaving phenotypic information embedded in drug–cell response matrices underutilized. Prior work has predominantly focused on IC50 regression or broad response prediction rather than explicit Super-Class MoA classification under strict leakage control. To address this gap, we propose a leakage-safe framework based on autoencoder-derived latent representations combined with a stacked neural–XGBoost classifier for stable and interpretable Super-Class MoA prediction.

Machine learning has been widely applied to drug mechanism and drug-response prediction across diverse biological settings. Early multilabel studies on the LISH-MoA dataset showed that neural networks outperform classical baselines in high-dimensional MoA prediction [1], while ensemble classifiers using engineered cardiac-tissue measurements improved mechanistic classification over single-feature approaches [2]. Deep perturbational models such as GPAR leveraged L1000 transcriptomics to surpass GSEA-style baselines [3], and split-network ensembles stabilized noisy gene-expression features for multiclass MoA tasks [4]. Generative models like Dr.VAE learned mechanistic latent spaces by jointly modeling perturbations and viability [5], whereas IC50 predictors based on autoencoders used compressed signatures to enhance cross-tissue generalization [6]. Probabilistic autoencoders incorporating dose–response uncertainty further improved IC50 consistency [7], with additional gains reported using multiomics autoencoders [8].

Domain-adaptation frameworks enabled cross-dataset drug-sensitivity transfer [9], while dual-VAE architectures integrating multiomics and drug features achieved robust regression and classification performance [10]. Interpretable models such as DrugCell mapped genotypes into pathway-structured networks for mechanistic prediction [11], and transfer-learning approaches like DADSP corrected cross-database feature misalignment to improve IC50 accuracy [12]. More recently, multimodal transformers trained on baseline and perturbational transcriptomes demonstrated improved generalization across unseen drugs and cell states [13], while contrastive models such as MoleculeSTM produced chemically meaningful embeddings beyond traditional fingerprints [14]. With deeper integration of omics and drug representations, multi-branch, graph-based, and ensemble pipelines have achieved improved predictions of drug sensitivity and synergy, particularly in low-data biomedical settings [15]–[17]. Despite these advances, leakage-safe Super-Class MoA classification using phenotypic IC50 data alone remains largely unexplored.

Model interpretability is critical for MoA prediction, as misclassification can mislead downstream biological analysis. While conventional metrics such as accuracy, macro-F1, and AUROC quantify predictive performance, they provide limited insight into model behavior. Although interpretability tools such as SHAP values, latent-space visualizations, and confusion-matrix analyses offer deeper understanding, they are rarely integrated into complete, leakage-controlled MoA prediction pipelines.

The aim of this study is to develop and evaluate a leakage-safe, phenotype-driven machine learning framework for predicting Super-Class drug mechanisms of action directly from large-scale IC50 cell-line sensitivity profiles. To this end, we introduce an embedding-driven pipeline incorporating standardized preprocessing, within-fold SMOTE, strict cross-validation, and stacked ensemble learning, supported by performance and interpretability analyses. The key contributions are outlined below:

- **Leakage-Safe Pipeline** – A strict evaluation setup with within-fold augmentation, standardized encodings, and stratified splits to prevent information leakage.
- **Stacked MoA Model** – An ensemble framework with unified training and scoring, enabling clear comparison of AUROC, F1, and label-wise performance trade-offs.
- **Explainable Diagnostics** – SHAP maps, latent-space clusters, reliability curves, and confusion-matrix analyses for mechanistic interpretability.
- **Operational Framework** – Calibration, monitoring, governance, and human-in-the-loop review to ensure scientific robustness and reproducibility.

The rest of the paper is organized as follows. Section II describes the dataset and model architecture. Section III presents the evaluation and interpretability analysis. Section IV concludes the paper.

II. METHODOLOGY

We introduce a leakage-safe framework for predicting drug Super-Class Mechanisms-of-Action (Super-Class MoA) from drug–cell line sensitivity profiles. The methodology integrates standardized preprocessing, autoencoder-based latent feature learning, and within-fold SMOTE to address class imbalance. A stacked neural–boosted ensemble was trained and evaluated under strict five-fold outer cross-validation. The framework is designed to ensure reproducibility and unbiased evaluation through fold-isolated preprocessing, representation learning, and model training.

A. Dataset Overview

The Super-Class MoA dataset consists of 28,700 drug–cell line sensitivity profiles across three expert-defined Super-Class MoA categories. Each sample corresponds to a unique drug–cell pair associated with a single IC50 value, validated drug and cell-line identifiers, and a Super-Class MoA label. The dataset was derived from the GDSC resource and contains 295 unique drugs and 969 cancer cell lines, with no missing IC50 values and no duplicate measurements. All IC50 values were log-transformed (LN_IC50). Super-Class MoA labels were curated from external pharmacological annotation sources and aligned with drug identifiers prior to modeling. After processing, the class distribution comprised 14,800 EPIGENETIC APOP METAB, 9,800 SIGNALING PATHWAYS, and 4,100 PROLIFERATION CONTROL samples are shown in Table I. This intrinsic imbalance reflects realistic pharmacological screening conditions and motivates the use of within-fold resampling during model training.

TABLE I. DISTRIBUTION OF LABELED SUPER-CLASS MOA

Super-Class MoA Class	Samples
EPIGENETIC APOP METAB	14800
SIGNALING PATHWAYS	9800
PROLIFERATION CONTROL	4100
Total	28700

B. Data Pre-processing

All preprocessing was done inside each outer cross-validation fold to strictly maintain a leakage-free setup. Each sample included a drug identifier, a cell-line identifier, and a continuous IC50 value. Drug and cell-line identifiers were not used as model features, so no label encoding was applied. The IC50 matrix was constructed by pivoting the dataset into a drug × cell-line structure, producing one numerical feature vector per drug. Continuous features were standardized using the training fold mean μ and standard deviation σ:

$$x_{std} = \frac{x - \mu}{\sigma} \tag{1}$$

where x is the raw feature value and μ, σ were computed only from the training portion of each fold. No missing IC50 values were present in the dataset, so no imputation was required. To address the imbalance among the three Super-Class MoA groups (14,800; 9,800; 4,100 samples), SMOTE was applied only to the training latent embeddings after the autoencoder produced the 32-dimensional representation. Synthetic minority points were generated using:

$$x_{new} = x_i + \delta \cdot (x_{nn} - x_i), \quad \delta \sim U(0,1) \tag{2}$$

where x_i denotes a minority-class sample and x_{nn} its nearest neighbor in feature space. The class imbalance motivating this resampling strategy and the effect of SMOTE on the training distribution are illustrated in Fig. 1.

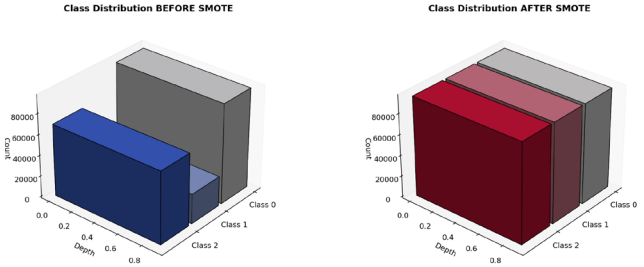


Fig. 1. Class Distribution Before and After SMOTE Oversampling

High correlation and redundancy among cell-line response features motivate the use of autoencoder-based representation learning for compact feature extraction. To strictly prevent data leakage, all preprocessing steps were performed within each outer cross-validation fold. The autoencoder was trained exclusively on standardized training data without resampling, and the frozen encoder was used to transform validation data. SMOTE was applied only to the latent embeddings of the training split after representation learning, while validation and test data remained completely untouched throughout training and evaluation.

C. Model Architecture

A symmetric dense autoencoder condenses standardized drug–cell characteristics into a 32-dimensional latent space. The embeddings are then fed into a neural network and an XGBoost classifier and consolidated through a logistic-regression meta-learner. Dimensionality 32 was chosen, as smaller bottlenecks (8, 16) lowered accuracy and larger bottlenecks (64) did not improve results. The neural model uses two dense layers (64 and 32 units, ReLU) with dropout 0.2, while the XGBoost model uses 300 trees, depth 6, and a 0.05 learning rate. Seed 42, Python 3.10, TensorFlow/Keras 2.x, and XGBoost 1.x are used for all experiments, and the model is trained for 100 epochs with a batch size of 64 using Adam (0.001). In this stacked setup, strict fold-level separation is maintained, and nonlinear feature learning is combined with boosted decision boundaries.

Autoencoder for Latent Representation Learning: Using a fully connected symmetric autoencoder, low-dimensional latent representations are learned from standardized drug–cell sensitivity features. As shown in Fig. 2, the encoder compresses the high-dimensional input through dense layers with ReLU activation and batch normalization, producing a 32-dimensional bottleneck vector that summarizes dominant drug-response variation, cell-line sensitivity patterns, and shared pharmacological structure. Reconstruction from this latent space suppresses noise and redundant correlations, yielding compact descriptors that improve Super-Class MoA discrimination. The decoder mirrors the encoder architecture and is trained by minimizing the mean-squared reconstruction error (MSE):

$$L_{rec} = \frac{1}{N}\sum_{i=1}^{N} \| x_i - \hat{x}_i \|_2^2 \qquad (3)$$

where x_i denotes the original feature vector and $\hat{x}_i$ its reconstruction. The autoencoder is trained strictly on the training split of each outer fold, and the encoder parameters are frozen before transforming validation samples, ensuring a fully leakage-free representation-learning process. This representation learning step assumes that dominant mechanistic variation can be embedded within a continuous latent space. While this assumption may not hold for all biological responses, the learned embeddings are not used in isolation; instead, downstream classification integrates a stacked neural–boosted ensemble to mitigate potential deviations from ideal latent continuity.

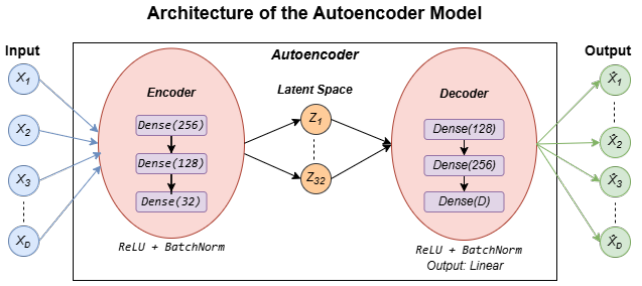


Fig. 2. Autoencoder architecture for latent feature extraction from drug–cell line profiles.

Stacked Neural–Boosted Ensemble Classifier: The module for classification becomes a stack of neural-boosted ensembles. This means that it will combine the predictive strengths of a neural classifier and a tree classifier operating on the 32 latent dimensions of the embeddings. As can be seen in Fig. 3, an autoencoder first compresses each drug–cell line feature vector into the compact representation $z \in \mathbb{R}^{32}$, which is then routed into two parallel base models. SMOTE is applied only to the training latent embeddings, after the autoencoder has generated the 32-D representations, ensuring that the autoencoder itself is trained solely on the original standardized data. The first branch turns out to be a shallow fully connected neural network that processes z through dense layers with ReLU activations, learning smooth nonlinear relationships and gradually represented mechanistic transitions trained into the latent space. Thus, the model is sensitive to subtle changes within the patterns of IC50 across related classes of drugs or similar responses from cell lines. The second part of the model employs an XGBoost classifier suitable for modeling abrupt decision boundaries that can build complicated interactions among the features. This is important because mechanisms for drug responses are threshold-like behaviors that neural networks cannot quite catch. Both models generate independent probability vectors over the three Super-Class MoA categories: $h_{\mathrm{NN}}(z)$ for the neural network and $h_{\mathrm{XGB}}(z)$ for the gradient-boosted trees. After concatenation, this is forwarded towards a logistic regression meta-classifier, which makes the final prediction as follows:

$$P(y \mid z) = \sigma\left(w_1 h_{NN}(z) + w_2 h_{XGB}(z) + b\right) \qquad (4)$$

Here, w_1 and w_2 are the learned weights for the neural and XGBoost branches, b is the bias term, and $\sigma(\cdot)$ maps the fused score to a probability. Enhanced stability, class separability, and interpretability of contributions from each branch are better achieved by implementing a stacked design that takes advantage of the smooth nonlinear patterns produced by neural networks with the sharp decision boundaries of boosted trees. This complementary stacking strategy reduces reliance on strict latent-space continuity by jointly modeling smooth latent trends and threshold-like decision boundaries, thereby improving robustness to heterogeneous or discontinuous mechanistic patterns.

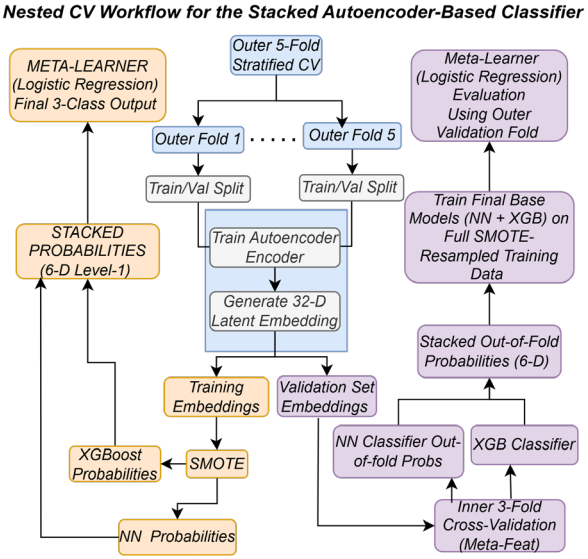


Fig. 3. Stacked neural–boosted ensemble classifier architecture.

Training and Optimization Setup: They all adopted an identical optimization regimen for similar behavioral prediction across folds. The autoencoder and neural classifier were trained using Adam (learning rate 1×10^{-4}), with a batch size of 64 and mean-squared reconstruction loss or categorical cross-entropy. Early stopping (patience 8) and 50 epochs resulted in the prevention of overfitting. With 300 estimators, depth 6, 0.05 learning rate, 0.8 subsample, and column-sample 0.8, XGBoost was also executed. The logistic-regression meta-learner was specified with L2 regularization and balanced class weights. All experiments were fixed by random seeds and run on an NVIDIA T4 GPU for reproducibility.

Overall Comparison: The proposed pipeline proved superior to all the stand-alone model baselines in terms of accuracy, clear class separation, and highly stable fold-wise behavior. The autoencoder produces substantially stronger latent representations, and the stacked neural-boosted ensemble is consistently better than the separate neural networks as well as XGBoost. Overall, the framework provides reliable and interpretable predictions for large-scale Super-Class MoA classification.

III. Result and Discussion

We present a comprehensive evaluation of the proposed Super-Class MoA prediction framework, which involves autoencoding for latent representation learning and stacking with neural–XGBoost. Performance was examined based on class-wise discrimination, cross-validation stability, and interpretability derived from SHAP-based explanations. Given the imbalanced nature of the dataset, macro-F1 and macro-AUROC are reported to provide fair and generalized performance measures. Fold behavior is analyzed through confusion matrices and ROC curves. Altogether, the assessments demonstrate that the two-stage pipeline is dependable, robust, and interpretable for large-scale Super-Class MoA prediction, addressing the need for scalable and explainable models in pharmacological screening.

Performance Stability Across Folds: Performance metrics were evaluated over the five outer folds of cross-validation. The model demonstrated stable performance with very low variance: accuracy 0.9182 ± 0.004, macro-F1 0.9118 ± 0.006, and macro-AUROC 0.9568 ± 0.003. These values represent the mean and standard deviation across folds and indicate consistent generalization with no fold-specific inflation.

A. Performance Stability Across Outer Cross-Validation

Model performance was evaluated under strict five-fold outer cross-validation to assess generalization robustness. Across folds, the proposed framework achieved a mean accuracy of 0.9182 ± 0.004, a mean macro-F1 score of 0.9118 ± 0.006, and a mean macro-AUROC of 0.9568 ± 0.003. The narrow standard deviations indicate highly consistent behavior across different train–validation splits, with no evidence of fold-specific inflation or instability.

Importantly, this stability is observed despite the substantial imbalance among Super-Class MoA categories, demonstrating that the combination of latent representation learning and within-fold resampling yields a robust decision function. Performance remained stable across the five outer folds, with low variance in accuracy, macro-F1, and macro-AUROC, confirming that the learned representations and classifier boundaries generalize reliably under repeated data partitioning.

B. Latent Space Structure and Class Separability

To examine whether the learned latent representations preserve discriminative mechanistic information, a t-SNE projection was generated from the 32-dimensional autoencoder embeddings. As illustrated in Fig. 4, the three Super-Class MoA categories form coherent and interpretable regions in the latent space. The EPIGENETIC APOP METAB class exhibits a compact and high-density cluster, suggesting strong internal consistency in its phenotypic response patterns. The SIGNALING PATHWAYS class forms a similarly cohesive structure with a well-defined boundary relative to the other classes. In contrast, PROLIFERATION CONTROL appears more dispersed and partially overlapping with neighboring regions. This behavior is consistent with the known biological heterogeneity of proliferation-related mechanisms and their overlap with epigenetic and signaling processes. Despite this natural overlap, the global geometry of the latent space demonstrates clear class-level organization, indicating that the autoencoder successfully captures dominant mechanistic variation while filtering out noise and redundant correlations. These findings support the suitability of the learned embeddings as inputs for downstream classification.

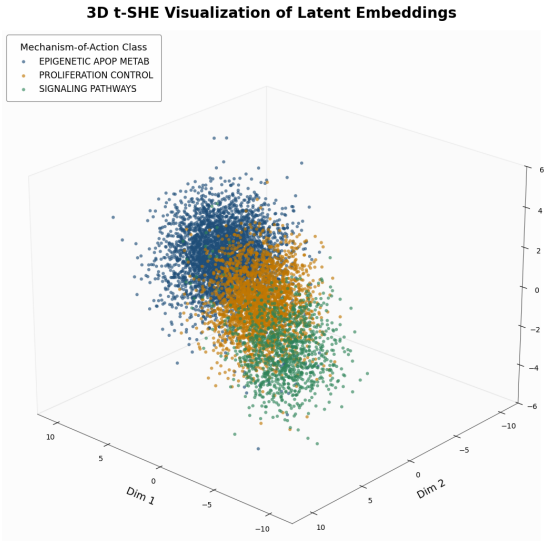


Fig. 4. Three-dimensional t-SNE visualization of learned latent embeddings.

C. Model Performance Across Outer Cross-Validation

The stacked neural–boosted ensemble achieved consistently strong discrimination across all Super-Class MoA categories. As shown in Fig. 5, the class-wise ROC curves yield AUROC values of 0.952 for EPIGENETIC APOP METAB, 0.918 for PROLIFERATION CONTROL, and 0.955 for SIGNALING PATHWAYS. The micro-average AUROC of 0.957 confirms robust threshold-independent ranking across the full dataset. To further characterize class-specific behavior under imbalance, precision, recall, and F1-scores were averaged across the five outer folds. EPIGENETIC APOP METAB achieved a precision of 0.923, recall of 0.904, and F1-score of 0.913, while SIGNALING PATHWAYS achieved a precision of 0.910, recall of 0.921, and F1-score of 0.915. PROLIFERATION CONTROL, despite being the smallest and most heterogeneous class, achieved a precision of 0.874, recall of 0.826, and F1-score of 0.847. These results demonstrate that the proposed framework maintains balanced performance across all classes and does not sacrifice minority-class sensitivity in favor of majority-class accuracy.

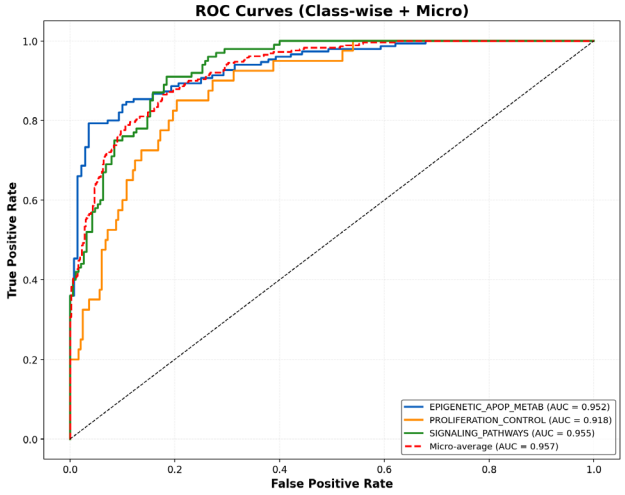


Fig. 5. Class-wise and micro-average ROC curves.

Cross-validation stability is further reflected in the aggregate performance metrics computed across the five outer folds. Fold-wise accuracy remained tightly bounded between 0.9030 and 0.9380, while macro-F1 ranged from 0.8990 to 0.9290, indicating consistent class-balanced performance. Macro-AUROC values were similarly stable, varying only between 0.9480 and 0.9660 across folds. This narrow variance demonstrates that neither the autoencoder's latent representations nor the ensemble classifier's decision boundaries are sensitive to data partitioning, even under substantial class imbalance. To quantify class-wise behavior under imbalance, precision, recall, and F1-scores were averaged across the five outer folds. EPIGENETIC APOP METAB achieved a precision of 0.923, recall of 0.904, and F1-score of 0.913; SIGNALING PATHWAYS achieved a precision of 0.910, recall of 0.921, and F1-score of 0.915; and PROLIFERATION CONTROL achieved a precision of 0.874, recall of 0.826, and F1-score of 0.847. These metrics corroborate the ROC and confusion-matrix analyses, demonstrating strong separability of the first two classes and competitive performance for the smaller, more heterogeneous PROLIFERATION CONTROL group.

D. Confusion Matrix Analysis

The average confusion matrix across the five outer folds is presented in Fig. 6 and provides further insight into class-wise decision behavior. The model correctly classifies the majority of samples across all Super-Class MoA categories, with the highest true-positive count observed for EPIGENETIC APOP METAB. SIGNALING PATHWAYS exhibits a balanced precision–recall profile, indicating stable recognition without excessive cross-class leakage. PROLIFERATION CONTROL shows moderate confusion with EPIGENETIC APOP METAB, reflecting overlapping phenotypic signatures and shared downstream biological effects. Importantly, most misclassifications occur near decision boundaries between mechanistically adjacent classes, rather than as random errors. This behavior aligns with known multi-pathway drug effects and further supports the biological plausibility of the learned decision boundaries.

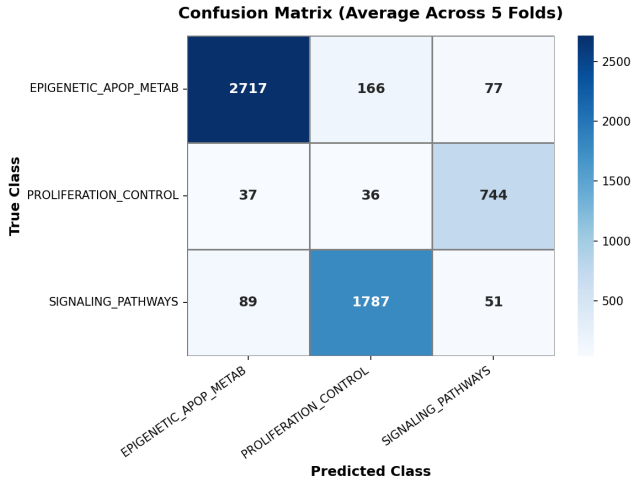


Fig. 6. Confusion matrix averaged across five cross-validation folds.

E. Feature Importance and Model Explainability

The interpretability of the model was evaluated through SHAP value analysis, which was performed solely on authentic validation samples to prevent synthetic SMOTE samples from affecting feature attribution. As illustrated in Fig. 7, SHAP analysis indicates that a small group of cell-line features, such as SKG-IIIa, RH-41, and SNU-16, are persistently strong contributors to Super-Class MoA prediction. Lower-ranked features exhibit minimal influence, implying that the ensemble relies on a focused and significant set of latent indicators rather than general associations. The SHAP distributions further demonstrate how individual features drive predictions toward specific Super-Class categories, providing insight into the direction and magnitude of feature impacts. Although these attributions do not imply direct causal relationships, they support the conclusion that the model utilizes biologically meaningful phenotypic patterns embedded within IC50 response profiles. Taken together, these analyses demonstrate that the proposed methodology achieves a balance between predictive performance and interpretability, which is essential for downstream pharmacological studies.

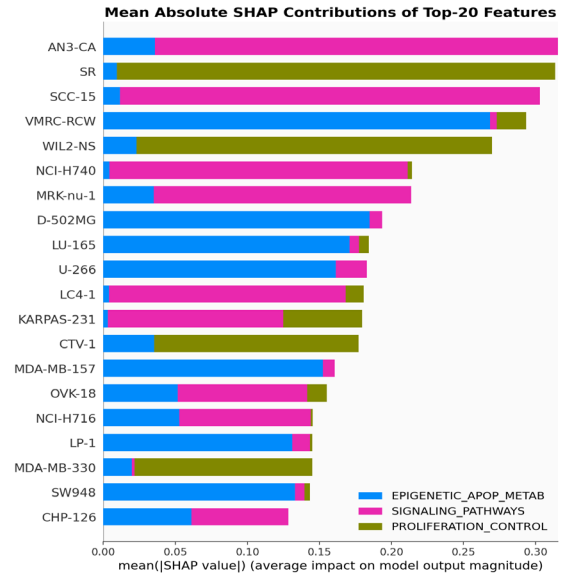


Fig. 7. Mean absolute SHAP contributions across Super-Class MoA.

F. Comparative Analysis

Currently available pharmacogenomic models primarily focus on IC50 regression tasks or broad multimodal drug-response prediction rather than explicit three-category Super-Class MoA classification under stringent leakage-control standards. Although VAE-based methods improve gene-expression representations, they mainly target binary sensitivity prediction rather than categorical MoA labels. Interpretable architectures such as DrugCell achieve strong correlations ($r = 0.80$) but remain regression-focused, while transfer-learning frameworks such as DADSP demonstrate reliable cross-database IC50 prediction (RMSE = 0.64) without addressing mechanistic class differentiation. Broader multi-task MoA systems [13] emphasize pathway inference rather than class-level separation, and graph-based ensembles such as GADRP [15] improve IC50 accuracy but are not structured for Super-Class MoA classification.

In contrast, the proposed framework integrates leakage-safe preprocessing, latent representation learning, and stacked neural–boosted fusion to directly address three-class Super-Class MoA prediction. Across five outer folds, the method achieves macro-F1 scores ranging from 0.899 to 0.929 and macro-AUROC values ranging from 0.948 to 0.966,

indicating strong class-wise separability and consistent generalization. These results position the proposed approach as a stable and interpretable alternative for Super-Class MoA prediction from phenotypic IC50 profiles.

TABLE II. COMPARATIVE SUMMARY WITH EXISTING METHODS

Ref.	Model Type / Domain	Metric	Result
[10]	VAE-based gene-expression MoA/sensitivity modeling	AUC / Accuracy	Up to 0.90
[11]	DrugCell interpretable deep model	Spearman ρ	0.80
[12]	DADSP transfer learning (IC50 regression)	RMSE / R^2	0.64 / 0.43
[13]	Transformer-based general drug-sensitivity model	Pearson r	≥0.85
[15]	GADRP graph-based multi-omics response predictor	RMSE (varies)	Best-in-class IC50 prediction
This work	Super-Class MoA classifier	Macro-F_1 / Macro-AUC	0.899–0.929 / 0.948–0.966

IV. Conclusion

This study presents a leakage-resistant, representation-based methodology for predicting Super-Class mechanism-of-action categories using IC50-derived sensitivity measurements, addressing a comparatively underexplored area of phenotypic MoA classification. Rigorous fold-specific preprocessing ensures statistical soundness through uniform encoding and intra-fold SMOTE rebalancing, thereby reducing the likelihood of significant leakage commonly encountered in pharmacogenomic analyses. The autoencoder serves as a central component by generating noise-reduced and feature-rich latent representations that capture the underlying structure of drug–cellular interactions while ensuring consistent feature transformations across validation folds. Experimental evaluation demonstrated that the hierarchical neural–boosted ensemble architecture systematically outperformed single-model approaches in terms of macro-F1, macro-AUROC, and cross-validation consistency. The synergistic behavior of the ensemble components was evident: the neural component effectively learned continuous nonlinear relationships within the latent space, whereas the XGBoost component modeled discrete, decision-oriented mechanistic patterns, with their integration enhancing model stability without inducing overfitting. Category-specific ROC analyses and confusion matrices indicated reliable discrimination among the three Super-Class MoA groups, despite substantial class distribution imbalance. SHAP-based explainability analysis further confirmed that model predictions were driven by a compact and informative subset of phenotypic features rather than spurious variations in IC50 values.

While the framework assumes that dominant mechanistic variation can be represented within a continuous latent space, this assumption may not hold uniformly across all biological contexts. The stacked ensemble design mitigates this limitation by combining smooth representation learning with tree-based decision boundaries, reducing reliance on strict latent continuity. Overall, the proposed pipeline provides a robust and interpretable approach for large-scale Super-Class MoA prediction using phenotypic sensitivity fingerprints alone. Future work will extend this framework toward finer-grained and multi-label MoA hierarchies, integrate multimodal drug descriptors, and explore transfer learning across pharmacogenomic datasets to further enhance early-stage drug screening and mechanism discovery.

References

[1] H. L. Gururaj, F. Flammini, H. a. C. Kumari, G. R. Puneeth, and B. R. S. Kumar, "Classification of drugs based on mechanism of action using machine learning techniques," *Discover Artificial Intelligence*, vol. 1, no. 1, Nov. 2021, doi: 10.1007/s44163-021-00012-2.

[2] R. Aghavali *et al.*, "Enhanced drug classification using machine learning with multiplexed cardiac contractility assays," *Pharmacological Research*, vol. 209, p. 107459, Oct. 2024, doi: 10.1016/j.phrs.2024.107459.

[3] S. Gao *et al.*, "Modeling drug mechanism of action with large scale gene-expression profiles using GPAR, an artificial intelligence platform," *BMC Bioinformatics*, vol. 22, no. 1, p. 17, Jan. 2021, doi: 10.1186/s12859-020-03915-6.

[4] R. U. Khan, "Prediction of Mechanism of Action Using Ensembled Deep Neural Networks Splits," *Revue D Intelligence Artificielle*, vol. 38, no. 5, pp. 1381–1389, Oct. 2024, doi: 10.18280/ria.380502.

[5] L. Rampášek, D. Hidru, P. Smirnov, B. Haibe-Kains, and A. Goldenberg, "Dr.VAE: improving drug response prediction via modeling of drug perturbation effects," *Bioinformatics*, vol. 35, no. 19, pp. 3743–3751, Mar. 2019, doi: 10.1093/bioinformatics/btz158.

[6] D. He, Q. Liu, Y. Wu, and L. Xie, "A context-aware deconfounding autoencoder for robust prediction of personalized clinical drug response from cell-line compound screening," *Nature Machine Intelligence*, vol. 4, no. 10, pp. 879–892, Oct. 2022, doi: 10.1038/s42256-022-00541-0.

[7] H. Dong, J. Xie, Z. Jing, and D. Ren, "Variational autoencoder for anti-cancer drug response prediction," *arXiv*, vol. 42, no. **7**, Sep. 2020, doi: 10.48550/arXiv.2008.09763.

[8] H. Wang *et al.*, "GADRP: graph convolutional networks and autoencoders for cancer drug response prediction," *Briefings in Bioinformatics*, vol. 24, no. 1, Dec. 2022, doi: 10.1093/bib/bbac501.

[9] X. Liao and B. Yu, "DvaDRP: a dual variational autoencoder based on big data for drug response prediction," in *Proc. 6th Int. Conf. Big-data Service and Intelligent Computation (BDSIC '24)*, New York, NY, USA: ACM, 2024, pp. 78–85, doi: 10.1145/3686540.3686551.

[10] Y. Li, D. M. Umbach, J. M. Krahn, I. Shats, X. Li, and L. Li, "Predicting tumor response to drugs based on gene-expression biomarkers of sensitivity learned from cancer cell lines," *BMC Genomics*, vol. 22, no. 1, p. 272, Apr. 2021, doi: 10.1186/s12864-021-07581-7.

[11] B. M. Kuenzi *et al.*, "Predicting drug response and synergy using a deep learning model of human cancer cells," *Cancer Cell*, vol. 38, no. 5, pp. 672-684.e6, Oct. 2020, doi: 10.1016/j.ccell.2020.09.014.

[12] W. Meng, X. Xu, Z. Xiao, L. Gao, and L. Yu, "Cancer drug sensitivity prediction based on deep transfer learning," *International Journal of Molecular Sciences*, vol. 26, no. 6, p. 2468, Mar. 2025, doi: 10.3390/ijms26062468.

[13] F. Carli *et al.*, "Learning and actioning general principles of cancer cell drug sensitivity," *Nature Communications*, vol. 16, no. 1, p. 1654, Feb. 2025, doi: 10.1038/s41467-025-56827-5.

[14] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Engineering Applications of Artificial Intelligence*, vol. 57, no. 9, 2022, doi: 10.48550/arXiv.2104.02395.

[15] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *Journal of King Saud University – Computer and Information Sciences*, vol. 35, no. 2, pp. 757–774, 2023, doi: 10.1016/j.jksuci.2023.01.014.

[16] W. Bakasa and S. Viriri, "Stacked ensemble deep learning for pancreas cancer classification using extreme gradient boosting," *Frontiers in Artificial Intelligence*, vol. 6, p. 1232640, Oct. 2023, doi: 10.3389/frai.2023.1232640.

[17] A. Ameen, N. Alganmi, and N. Bajnaid, "Stacked Deep Learning Ensemble for Multiomics Cancer Type Classification: Development and Validation Study," *JMIR Bioinformatics and Biotechnology*, vol. 6, p. e70709, Jun. 2025, doi: 10.2196/70709.