PNNL-38864

Finch: Toxicity Dose Response Curve Prediction of Chemical Compounds and Mixtures

October 2025

Abdullah Shouaib
John Zapanta
Sean P. Davern
Samuel Dixon
Zachary R. Stromberg
Becky Hess
Sydney Schwartz
C Mark Maupin

U.S. DEPARTMENT of ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-38864

Finch: Toxicity Dose Response Curve Prediction of Chemical Compounds and Mixtures

October 2025

Abdullah Shouaib
John Zapanta
Sean P. Davern
Samuel Dixon
Zachary R. Stromberg
Becky Hess
Sydney Schwartz
C Mark Maupin

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

A paradigm shift in chemical risk assessment is emphasizing mixture testing over single compound analysis, eliminating animal testing, and adopting advanced modeling approaches to understand mixture activity profiles. However, existing computational models largely focus on single chemicals, with few effective solutions for modeling complex mixtures that account for synergistic or antagonistic effects and multiple Modes of Action (MoA). Conventional methods like concentration addition (CA) and independent action (IA) are insufficient for this task as they are designed for simplistic interactions and struggle to account for the dynamic and multifaceted nature of chemical mixtures, such as overlapping MoA and non-linear interactions. Finch offers a novel approach utilizing deep learning (DL) embeddings and multi-task quantitative structure-activity relationship (QSAR) models to improve chemical exposure prediction. By leveraging molecular descriptors, physiochemical properties, and large language model (LLM) embeddings from SMILES inputs, Finch preserves critical information in a latent space thereby enhancing predictive accuracy. The multi-task learning aspect of Finch is highly advantageous, as it simultaneously optimizes multiple loss functions, leveraging all available data across tasks to develop generalized representations that effectively capture complex ingredient interactions within mixtures.

Acknowledgments

This research was supported by the National Security Mission Seed Investment, under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830. A portion of the research was performed using resources available through Research Computing at Pacific Northwest National Laboratory (PNNL). During the preparation of this report, the authors used OpenAI's gpt 4.o for the purpose of assisting in writing. The authors have reviewed and edited the output and take full responsibility for the content of this report.

Acronyms and Abbreviations

AI	Artificial Intelligence
CA	Concentration Addition
DL	Deep Learning
FMD	Formula Molecular Descriptors
FME	Formula Molecular Embeddings
IA	Independent Action
LLM	Large Language Model
MoA	Mode of Action
MD	Molecular Descriptors
ME	Molecular Embeddings
ML	Machine Learning
OPFRS	Organophosphate Flame Retardants
PFAS	Polyfluoroalkyl Compounds
RF	Random Forest
QSAR	Quantitative Structure Activity Relationship
RCDK	R's Chemistry Development Kit
ROC-AUC	Receiver Operating Characteristic Area under the Curve
SMILES	Simplified Molecular Input Line Entry System
2,4-DCP	2,4-dichlorophenol
2,5-DCP	2,5-dichlorophenol
BP-3	Benzophenone-3
bPB	Butyl paraben
Cd	Cadmium chloride hydrate
Co	Cobalt chloride
Cu	Cupric sulfate
EHDPHP	2-Ethylhexyl diphenyl phosphate
Hg	Methylmercury chloride
Ni	Nickel dichloride
Pb	Lead chloride
PFHxS	Perfluorohexanesulfonic acid
PFNA	Perfluorononanoic acid
PFOA	Perfluorooctanoic acid
pPB	Propyl paraben
Sb	Antimony(III) chloride
Se	Sodium selenite
TBOEP	Tris(2-butoxyethyl) phosphate

TCPP	Tris(1-chloro-2-propyl)phosphate
TCS	Triclosan
TDCPP	Tris(1,3-dichloropropyl) phosphate
TEHP	Tri (2-ethylhexyl)phosphate
TPhP	Triphenyl phosphate
Zn	Zinc sulfate heptahydrate

Contents

Figures

Tables

1.0 Introduction

The accurate prediction of dose-response behavior in chemical mixtures is a central challenge within the field of toxicology. Unlike single-agent exposures, mixtures often exhibit complex, non-linear interactions such as antagonism, synergism, or potentiation, which complicate efforts to anticipate their biological impacts. [1-5] Addressing this complexity is essential for safeguarding public health, promoting environmental safety, and providing reliable guidance for regulatory decision-making and product development.

Traditional toxicological testing approaches, relying extensively on *in vivo* and *in vitro* experiments, face significant limitations when applied to chemical mixtures. These methods are both resource-intensive and constrained by the sheer number of possible combinations that require assessment. Furthermore, experimental techniques designed to quantify multi-component interactions often demand advanced robotics and specialized equipment, thereby adding cost and limiting accessibility for many research laboratories. Consequently, the development of novel, scalable, and cost-effective alternatives has become an urgent priority.

Recent advancements in artificial intelligence (AI) and machine learning (ML) provide a promising framework for addressing this challenge. [6-14] These methodologies offer the ability to integrate large-scale toxicological datasets and model complex chemical interactions efficiently. Specifically, deep learning (DL) architectures have demonstrated significant capabilities in identifying hierarchical features and capturing nuanced relationships among toxicological endpoints and chemical properties. [15-17] This computational approach allows for rapid, reliable predictions of mixture toxicity, reducing reliance on traditional experiments and enabling novel insights into mixture behaviors that would otherwise be difficult to measure.

Despite these technological advances, predictive modeling for mixture toxicity still faces critical obstacles, most notably data scarcity. For chemical mixtures, datasets are often insufficient due to variability in composition, concentration, and endpoint measurements. Furthermore, the combinatorial complexity inherent to multi-component mixtures far exceeds the data generation capacity of traditional experimental approaches. Addressing these issues requires innovative strategies capable of leveraging existing data effectively while adapting to the diverse challenges posed by mixtures.

In this context, the emergence of transfer learning represents a significant step forward. This approach enables models trained on single-agent toxicity data to be adapted for multi-component formulations, reducing the requirement for mixture-specific datasets. [18-20] By leveraging pre-established representations and learned parameters from single-chemical datasets, transfer learning not only mitigates data limitations but also improves predictive accuracy in modeling complex interactions. Implementing transfer learning into toxicity prediction frameworks offers opportunities to expedite the identification of hazardous combinations, guide experimental protocols, and inform risk assessments with greater precision and efficiency.

2.0 Materials and Methods

2.1 Data Collection and Processing

2.1.1 Cytotoxicity of Mixtures Data

Cytotoxicity data for individual chemicals and mixtures were sourced from a previously established dataset linked to the HBM4EU project, encompassing 24 compounds and 39 mixtures. [8, 12] The selected compounds included 9 heavy metals, 6 organophosphate flame retardants (OPFRs), 3 polyfluoroalkyl substances (PFAS), and 6 phenols. This dataset provided molecular structure information via Simplified Molecular Input Line Entry System (SMILES), mixture compositions, and toxicity measurements across varying concentrations. To prepare the data for model training, cytotoxicity values were normalized to a range of 0 to 1 using min-max scaling, with a maximum cap of 100%. Concentration values were converted to molar units to enhance numerical stability during model development.

2.1.2 PubChem Bioassay Data

The primary dataset was sourced from PubChem and comprised bioassay data on the toxicity of 9,524 compounds in HepG2 cell lines at exposure times of 24 hours and 40 hours. The dataset included SMILES strings, activity outcomes, and assay parameters, with compounds classified as "Active" (toxic) or "Inactive" (non-toxic) based on the "PUBCHEM_ACTIVITY_OUTCOME" field. Canonicalized SMILES representations were extracted for model input, and binary labels were assigned (1 for toxic and 0 for non-toxic compounds). The data was stratified into training (80%) and validation (20%) subsets, ensuring consistent class distributions across splits.

2.2 ChemBERTa-2 Fine-Tuning

The ChemBERTa-2 model, accessed via Hugging Face, was utilized to perform binary classification of compound cytotoxicity. All workflows including data preprocessing, model fine-tuning, and embedding extraction were implemented in Python (v3.10) using libraries such as pandas (v1.4), numpy (v1.24), PyTorch (v2.5), and Hugging Face Transformers (v4.33). Chemical structures, represented as SMILES strings, were tokenized using ChemBERTa-2's custom tokenizer to generate token sequences formatted for the model, including special tokens ([CLS] and [SEP]) suitable for its BERT-like architecture.

During fine-tuning, the PubChem bioassay data was processed, where each encoded input chemical leveraged the [CLS] token to pass through a classification head, yielding probabilities for cytotoxic versus non-cytotoxic outcomes. After training, embeddings (384-dimensional representations generated from the [CLS] token) were extracted using the Hugging Face pipeline and stored for downstream analysis, including clustering and additional machine learning applications.

Model training was conducted on a DGX node equipped with 8 2080-ti GPUs to accelerate computation, using 200 epochs, a batch size of 64, the AdamW optimizer with a learning rate of 10^{-5}, a linear scheduler with 500 warmup steps, and a weight decay of 0.01. Performance evaluation was conducted on a held-out validation set (20% of the data) using metrics including accuracy, F1-score, and Receiver Operating Characteristic Area Under the Curve (ROC-AUC).

2.3 Molecular Descriptor Generation

Molecular descriptors for individual compounds from the previously obtained dataset [8, 12] were generated using the Chemistry Development Kit (RCDK) in R [21]. These descriptors captured a range of physicochemical properties, including topological, geometric, and electronic features. After filtering out descriptors with missing values (NA) or no variability, a total of 103 descriptors remained, which were normalized using min-max scaling.

For chemical mixtures, formula molecular descriptors (FMDs) were computed as weighted sums of the individual molecular descriptors (MD), with weights based on the mole fractions of each compound within the mixture:

$$FMD_i = \sum_{n}^{n_{max}} \left(x_n \times MD_{i,n}\right)$$

Here, FMD_i represents the ith molecular descriptor for a mixture, χ_n is the mole fraction of compound n, and $MD_{i,n}$ is the ith descriptor of the nth individual compound. This methodology produces a composite descriptor vector that captures the aggregated molecular properties of the mixture. The composite FMDs were utilized as input features for training random forest regression models, enabling predictive analysis based on the molecular representations of both individual compounds and mixtures.

2.4 Molecular Embedding Extraction

Molecular embeddings derived from the pre-trained and fine-tuned ChemBERTa models were utilized to capture chemical information relevant to toxicity prediction. These embeddings served as robust molecular representations for downstream machine learning tasks, including concentration-dependent toxicity modeling.

To extract embeddings for individual compounds, SMILES strings were processed through both versions of the ChemBERTa model, with the final hidden layer output corresponding to the [CLS] token captured as 384-dimensional vectors. This process was efficiently executed using PyTorch's no-gradient context for batch processing. These embeddings encapsulate chemical information learned during fine-tuning and were subsequently used as descriptors for predictive modeling.

For mixture toxicity modeling, a formula molecular embedding (FME) was calculated by extending the concept used for molecular descriptors (FMDs). FME was computed as a weighted sum of the individual compound embeddings, proportional to each compound's mole fraction in the mixture:

$$FME_i = \sum_{n}^{n_{max}} \left(x_n \times [CLS]_{i,n}\right)$$

Here, FME_i represents the ith formula molecular embedding for a mixture, χ_n is the mole fraction of compound n, and $[CLS]_{i,n}$ is the ith element of the [CLS] embedding for the nth compound. This approach produces a single composite embedding vector that encapsulates the aggregated molecular information of the mixture. These composite embeddings were subsequently employed as input features for training random forest regression models to predict mixture toxicity.

2.5 Random Forest Models for Concentration-Dependent Toxicity Prediction

Three Random Forest models were developed to evaluate the effectiveness of molecular descriptors (MDs and FMDs) versus embeddings from pre-trained and fine-tuned ChemBERTa models (MEs and FMEs) in predicting concentration-dependent cytotoxicity responses. This methodology leveraged the detailed chemical information captured by MDs and MEs, while utilizing machine learning models to account for non-linear relationships between chemical structure and toxicity across concentration levels.

The Random Forest models were implemented using scikit-learn (v1.7.2) with comprehensive hyperparameter optimization for robust performance. Input feature vectors for the MD/FMD-based models included 103 molecular descriptors along with compound concentration values, while the ME/FME-based models utilized 384-dimensional ChemBERTa embeddings combined with concentration attributes. The target variable was normalized cytotoxicity, ranging from 0 (no cytotoxicity) to 1 (complete cytotoxicity).

Hyperparameters such as the number of estimators (trees) [100, 200, 300, 400], maximum tree depths [None, 10, 20, 30, 40], and split criteria ['squared_error', 'absolute_error'] were optimized via extensive grid search. Optimal parameters included 200 estimators and a maximum depth of 10 for the MD-based model, and 100 estimators with no depth restriction for the ME-based model. Five-fold cross-validation, with stratified sampling by compounds, was employed to prevent overfitting, ensuring that all concentrations of a compound were retained within the same fold. This stratification was essential to assess model generalizability across unseen chemical structures.

3.0 Results

3.1 ChemBERTa-2 Fine-Tuning

The ChemBERTa-2 model was fine-tuned on HepG2 toxicity data from PubChem to develop a domain-specific model capable of generating toxicology-tailored embeddings. To evaluate the fine-tuned model, key metrics including accuracy, Receiver Operating Characteristic Area Under the Curve (ROC-AUC), and F1-score were employed, ensuring a robust assessment of the classifier's performance. Accuracy provided a straightforward measure of correct predictions relative to total predictions, while ROC-AUC evaluated the model's sensitivity and specificity across varying thresholds, reflecting its ability to discriminate between toxic and non-toxic compounds. The F1-score, combining precision and recall, was particularly valuable for analyzing class imbalances by balancing false positives and false negatives. Collectively, these metrics offered a comprehensive view of the model's strengths and limitations without over-reliance on a single measurement.

Table 1. Fine-Tuning of ChemBERTa-2 on PubChem Data.

Model	Accuracy	ROC-AUC	F1-Score
ChemBERTa Classifier	0.89	0.70	0.41

Results from the fine-tuning process are shown in Table 1, where the ChemBERTa-2 classifier achieved an accuracy of 0.89, indicating a high rate of correct predictions. The ROC-AUC score of 0.70 suggests moderate ability to differentiate between toxic and non-toxic compounds but reveals limitations in capturing subtleties at varying thresholds. Additionally, the F1-score of 0.41 highlights challenges, primarily due to high false negatives, reflecting low recall on toxic predictions.

To further assess performance, a confusion matrix analysis was conducted (Figure 1), providing a detailed breakdown of the model's predictions relative to ground truth. The analysis revealed strong performance in classifying non-toxic compounds correctly (0.93 accuracy), with a low misclassification rate (0.07). However, the model struggled to accurately identify toxic compounds, achieving only 0.47 correctness in this category. These limitations likely stem from

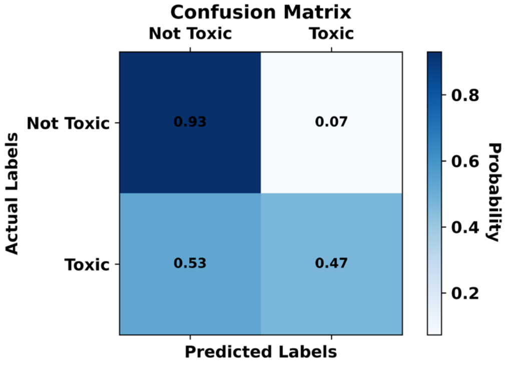

Figure 1. Confusion matrix for the ChemBERTa-based classification model predicting HepG2 toxicity outcome.

the inherent noise within the HepG2 PubChem dataset, which is characterized by high-throughput screening issues, such as variability in dose-response curves. This noise is attributed to small assay volumes in 1536-well plates, leading to inconsistencies in curve categorization and, consequently, impacting the fine-tuning process. While the overall metrics demonstrate promising potential for ChemBERTa-2 in toxicology applications, the variability in data quality presents challenges that must be addressed to further enhance model performance.

3.2 Model Validation Results

Three Random Forest (RF) regression models were trained to predict dose-response curves for chemical compounds and mixtures, utilizing distinct molecular representations: a molecular descriptor-based RF model (MD RF), a pre-trained molecular embedding RF model (pre-trained ME RF), and a fine-tuned molecular embedding RF model (fine-tuned ME RF). Model validation was performed using 5-fold cross-validation on the training dataset (n=554), a standard technique to mitigate overfitting and provide a robust evaluation of model performance.

The accuracy of the models was assessed using three metrics: R-squared (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), as outlined in Table 2. R^2 measures the global fit of the model, MAE provides an intuitive measure of average prediction error, and RMSE prioritizes sensitivity to larger deviations. Utilizing all three metrics ensures a comprehensive evaluation and allows for meaningful comparison across models. Validation results indicated strong performance across all models, with both the MD RF and fine-tuned ME RF models achieving an R^2 of 0.89 ± 0.05, while the pre-trained ME RF model performed comparably with an R^2 of 0.90 ± 0.04. These values demonstrate the models' ability to explain a substantial proportion of variance in the target variable. Predictive accuracy, reflected in MAE, was similar across models, with low values of 0.066 ± 0.02 (MD RF), 0.065 ± 0.02 (pre-trained ME RF), and 0.068 ± 0.02 (fine-tuned ME RF). Additionally, RMSE values were consistent, with 0.14 ± 0.03 for both MD RF and fine-tuned ME RF models, and 0.13 ± 0.03 for the pre-trained ME RF model, emphasizing the models' ability to limit larger prediction errors. Overall, the results indicate that all three RF models effectively capture and predict dose-response characteristics with comparable performance, demonstrating their reliability in toxicology modeling.

Table 2. Kfold Cross Validation Results on Train Data Set (n=554).

Model	R^2	MAE[1]	RMSE[1]
MD RF	0.89 ± 0.05	0.066 ± 0.02	0.14 ± 0.03
Pre-trained ME RF	**0.90 ± 0.04**	**0.065 ± 0.02**	**0.13 ± 0.03**
Fine-tuned ME RF	0.89 ± 0.05	0.068 ± 0.02	0.14 ± 0.03

[1] Units are Cytotoxicity.

3.3 Dose Response Curve Predictions

The performance of the three Random Forest models (MD RF, pre-trained ME RF, and fine-tuned ME RF) was evaluated on the test set (20% hold-out data) using R^2, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), as detailed in Table 3. Results demonstrated that all models achieved high accuracy, with minimal differences observed between the input feature types (MDs, pre-trained MEs, and fine-tuned MEs). Plots of predicted values against ground truth (Figure 2) revealed strong performance at the extreme values, while errors were more pronounced in the central region, likely due to the steeper slope changes in this range. Further evaluation of dose-response curves for individual compounds (Figure 3) and mixtures (Figure 4) provided additional insights into model strengths and weaknesses. Analysis of individual compound predictions indicated that pre-trained ME RF and fine-tuned ME RF models demonstrated marginal improvements

Table 3. Random Forest Model Metrics on Test Data Set (n=139).

Model	R^2	MAE[1]	RMSE[1]
MD RF	0.90	0.056	0.12
Pre-trained ME RF	**0.92**	**0.046**	**0.11**
Fine-tuned ME RF	0.91	0.052	0.12

[1] Units are Cytotoxicity.

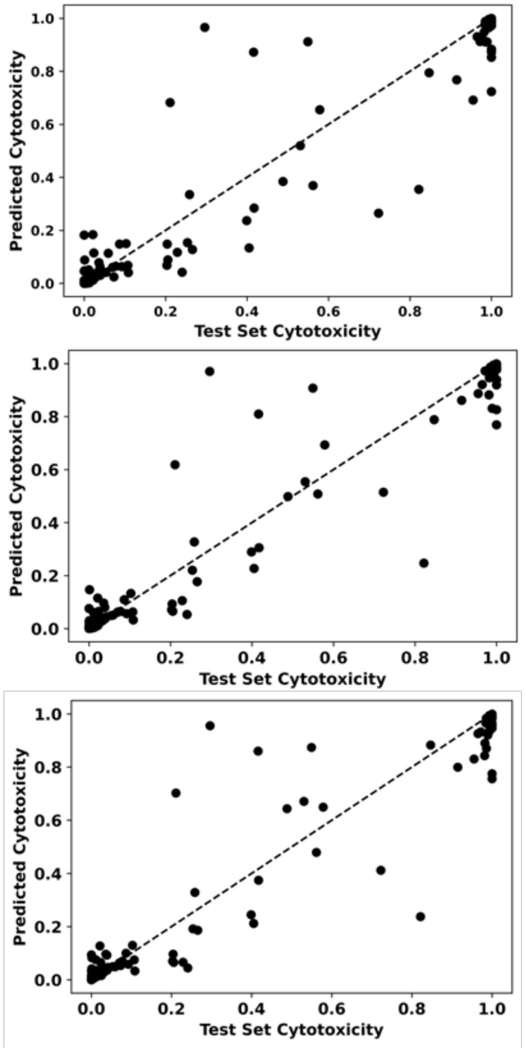

Figure 2. Test set ground truth against (Top) MD RF model (Middle) pre-trained ME RF model, and (Bottom) fine-tuned ME RF model.

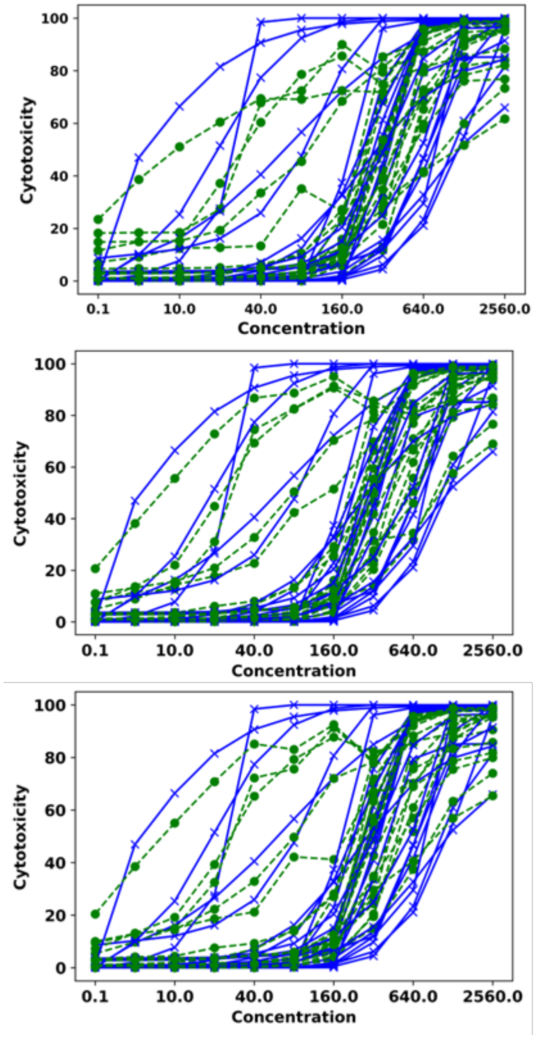

Figure 3. Individual compound dose response curve for (Top) MD RF, (Middle) pre-trained ME RF, and (Bottom) fine-tuned ME RF where blue (χ) is ground truth and the green (●) are predicted values.

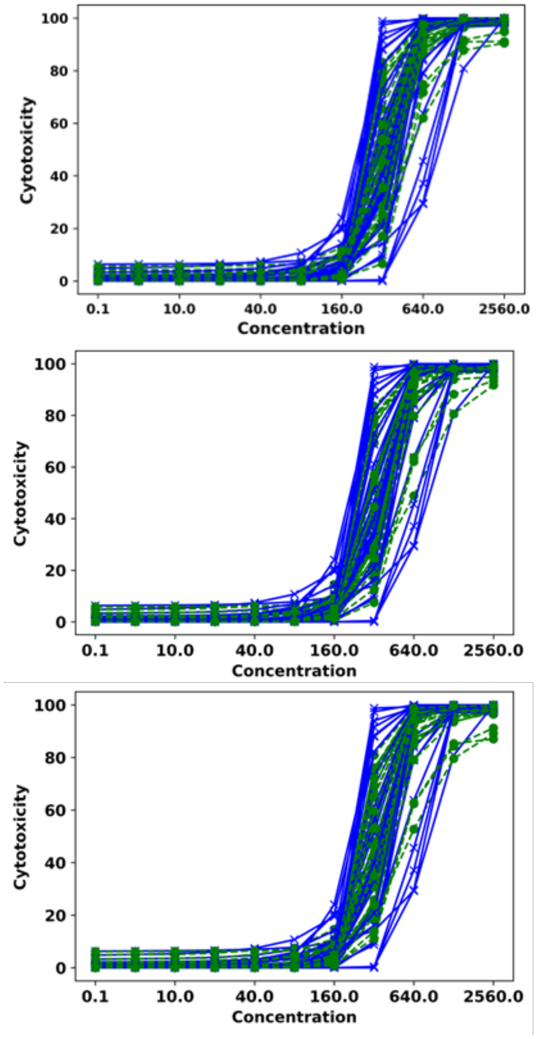

Figure 4. Mixture dose response curve for (Top) MD RF, (Middle) ME RF, and (Bottom) fine-tuned ME RF where blue (x) is ground truth and the green (●) are predicted values.

4.0 Discussion

Our analysis of molecular representation methods for liver toxicity prediction demonstrated comparable performance across approaches, with pre-trained ChemBERTa embeddings combined with Random Forest regression slightly outperforming fine-tuned ChemBERTa embeddings and traditional RCDK molecular descriptors. Specifically, the pre-trained embeddings achieved an R^2 of 0.92, compared to 0.91 for fine-tuned embeddings and 0.90 for RCDK descriptors. This result underscores the capability of pre-trained transformer-based embeddings to inherently capture chemical patterns relevant to toxicological outcomes, without requiring endpoint-specific fine-tuning. The transformer architecture in ChemBERTa effectively learns contextual chemical substructures and interactions, yielding robust performance comparable to traditional, descriptor-based methods reliant on predefined chemical features.

Fine-tuning ChemBERTa on HepG2 toxicity data provided only marginal performance improvement beyond pre-trained embeddings, likely due to limitations in the fine-tuning dataset. The dataset, originating from high-throughput microwell assays, introduced variability and noise due to experimental conditions such as small cell counts per well, compromising the training signal quality. This observation highlights the importance of high-quality and biologically representative data for effective model fine-tuning. Moreover, it suggests that leveraging pre-trained embeddings directly is a viable strategy when endpoint-specific datasets are of limited quality.

A key innovation in this study was the prediction of mixture toxicity using weighted molecular embeddings. By combining individual embeddings weighted by their mole fractions, we successfully generated representations of mixtures capable of capturing their toxicity profiles with high accuracy, evidenced by strong R^2 scores and low prediction errors. This approach provides a flexible alternative to traditional mixture prediction methods that often rely on extensive experimental data or simplistic models like concentration addition or independent action. The embedding-based framework allows predictions for mixtures of compounds with available individual embeddings, without requiring explicit mixture training data—a feature highly valuable for real-world risk assessment scenarios involving complex chemical exposures.

Another significant advancement was the ability to predict detailed dose-response curves for individual compounds and mixtures, moving beyond binary classification models that simply categorize compounds as toxic or non-toxic. By modeling toxicity as a function of concentration, our approach provides a more nuanced assessment aligned with the dose-dependent nature of toxicity. This capability enables the calculation of essential toxicological parameters such as EC10 and EC50 values, critical for regulatory decision-making and risk assessment. The close agreement observed between predicted and experimental dose-response curves further validates the robustness of this approach across diverse scenarios.

Despite these promising results, several limitations must be acknowledged. The current approach assumes additivity in the embedding space for mixture toxicity prediction, which may not fully account for potential synergistic or antagonistic interactions in complex mixtures. Additionally, training data derived from HepG2 cell line experiments may not fully represent the biological complexity of liver toxicity *in vivo*, including metabolic activation or detoxification mechanisms. The imbalance in the fine-tuning dataset, with a larger proportion of non-toxic compounds, could bias the model toward predicting non-toxicity. Furthermore, while embeddings capture rich chemical-toxicity relationships, the lack of interpretability for individual embedding dimensions limits insights into the underlying mechanisms of toxicity.

Future efforts should focus on addressing these challenges through several avenues. Incorporating metabolic activation data could enhance the physiological relevance of predictions, particularly for compounds requiring metabolic processing to exert toxic effects. Expanding the approach to additional cell types and toxicity endpoints would broaden its application beyond liver toxicity. Moreover, developing advanced mixture toxicity models that account for synergistic and antagonistic effects could further improve predictions for complex exposures. These enhancements, combined with efforts to improve dataset quality and diversity, will strengthen the utility and applicability of embedding-based methodologies for toxicological research and regulatory frameworks.

5.0 Conclusion

This study introduces a transformative approach to predictive toxicology by leveraging ChemBERTa-based molecular representations for liver toxicity prediction. The findings reveal that pre-trained ChemBERTa embeddings paired with Random Forest regression outperform fine-tuned embeddings and traditional descriptor-based methods. This result demonstrates the inherent capacity of transformer-based architectures to capture chemical features relevant to toxicological outcomes without endpoint-specific fine-tuning. Furthermore, the use of weighted molecular embeddings for mixture toxicity prediction represents a viable methodology for the characterization of complex mixture toxicity profiles. This approach addresses key challenges in real-world risk assessment and regulatory applications where exposure to chemical mixtures is more common than individual compounds. Additionally, the ability to predict complete concentration-response curves adds significant benefits such as allowing dose-dependent hazard evaluation and the calculation of critical toxicological parameters like EC10 and EC50 values.

6.0 References

1. Sarigiannis, D.A. and U. Hansen, *Considering the cumulative risk of mixtures of chemicals – A challenge for policy makers.* Environmental Health, 2012. **11**(1): p. S18.
2. Shaw, I.C., *Chemical residues, food additives and natural toxicants in food – the cocktail effect.* International Journal of Food Science and Technology, 2014. **49**(10): p. 2149-2157.
3. Wang, N., et al., *Prediction of the joint action of binary mixtures based on characteristic parameter k·ECx from concentration-response curves.* Ecotoxicology and Environmental Safety, 2021. **215**: p. 112155.
4. Cedergreen, N., *Quantifying synergy: a systematic review of mixture toxicity studies within environmental toxicology.* PLoS One, 2014. **9**(5): p. e96580.
5. Elcombe, C.S., E.N. P., and M. Bellingham, *Critical review and analysis of literature on low dose exposure to chemical mixtures in mammalian in vivo systems.* Critical Reviews in Toxicology, 2022. **52**(3): p. 221-238.
6. Wang, T., et al., *Prediction of the Toxicity of Binary Mixtures by QSAR Approach Using the Hypothetical Descriptors.* International Journal of Molecular Sciences, 2018. **19**(11): p. 3423.
7. Zhang, F., et al., *Machine learning-driven QSAR models for predicting the mixture toxicity of nanoparticles.* Environ Int, 2023. **177**: p. 108025.
8. Kim, J., M. Seo, and M. Na, *MRA Toolbox v. 1.0: a web-based toolbox for predicting mixture toxicity of chemical substances in chemical products.* Scientific Reports, 2022. **12**(1): p. 8880.
9. Chatterjee, M. and K. Roy, *Predictive binary mixture toxicity modeling of fluoroquinolones (FQs) and the projection of toxicity of hypothetical binary FQ mixtures: a combination of 2D-QSAR and machine-learning approaches.* Environmental Science: Processes & Impacts, 2024. **26**(1): p. 105-118.
10. Chatterjee, M. and K. Roy, *Prediction of aquatic toxicity of chemical mixtures by the QSAR approach using 2D structural descriptors.* Journal of Hazardous Materials, 2021. **408**: p. 124936.
11. Abbod, M. and A. Mohammad, *Combined interaction of fungicides binary mixtures: experimental study and machine learning-driven QSAR modeling.* Scientific Reports, 2024. **14**(1): p. 12700.
12. Kim, S., et al., *In Vitro Toxicity Screening of Fifty Complex Mixtures in HepG2 Cells.* Toxics, 2024. **12**(126): p. 1-12.
13. Ge, H., et al., *Integrative Assessment of Mixture Toxicity of Three Ionic Liquids on Acetylcholinesterase Using a Progressive Approach from 1D Point, 2D Curve, to 3D Surface.* Int J Mol Sci., 2019. **20**(21): p. 5330.
14. Tropsha, A., et al., *Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR.* Nature Reviews Drug Discovery, 2024. **23**(2): p. 141-155.
15. Sabando, M.V., et al., *Using molecular embeddings in QSAR modeling: does it make a difference?* Briefings in Bioinformatics, 2021. **23**(1).
16. Colby, S.M., et al., *Deep Learning to Generate in Silico Chemical Property Libraries and Candidate Molecules for Small Molecule Identification in Complex Samples.* Analytical Chemistry, 2020. **92**(2): p. 1720-1729.
17. Chen, J.-H. and Y.J. Tseng, *A general optimization protocol for molecular property prediction using a deep learning network.* Briefings in Bioinformatics, 2021. **23**(1).
18. Singh, S.P., *Transfer of learning by composing solutions of elemental sequential tasks.* Machine Learning, 1992. **8**(3): p. 323-339.
19. Simões, R.S., et al., *Transfer and Multi-task Learning in QSAR Modeling: Advances and Challenges.* Front Pharmacol, 2018. **9**: p. 74.

20. Pratt, L. and B. Jennings, *A Survey of Transfer Between Connectionist Networks.* Connection Science, 1996. **8**(2): p. 163-184.

21. Steinbeck, C., et al., *The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo and Bioinformatics.* J. Chem. Inf. Comput. Sci., 2003. **43**: p. 493-500.

Pacific Northwest
National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov