

Supporting Information

Rapid Prediction of Chemical Ecotoxicity Through Genetic Algorithm Optimized Neural Network Models

Ping Hou^{1,2}, Bu Zhao^{1,2}, Olivier Jolliet³, Ji Zhu⁴, Peng Wang⁵, and Ming Xu^{1,7*}

¹ School for Environment and Sustainability, University of Michigan, 440 Church St., Ann Arbor, MI, 48109, USA.

² Michigan Institute for Computational Discovery & Engineering, University of Michigan, 500 Church St., Ann Arbor, MI, 48109, USA.

³ Environmental Health Sciences, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI, 48109, USA.

⁴ Department of Statistics, University of Michigan, 1085 South University, Ann Arbor, MI, 48109, USA.

⁵ Department of Electrical and Computer Engineering, University of Kentucky, 512 Administration Drive, Lexington, KY, 40506, USA.


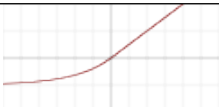
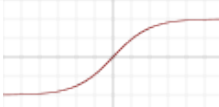




⁶ Department of Mechanical Engineering, University of Kentucky, 151 Ralph G. Anderson Building, Lexington, KY, 40506, USA.

⁷ Department of Civil and Environmental Engineering, University of Michigan, 2350 Hayward Street, Ann Arbor, MI, 48109, USA.

*Correspondence to: mingxu@umich.edu.

This Supporting Information includes 3 pages, 2 figures and 1 table.

Table S1. Commonly used activation functions.

Name	Input output relation	Graphic representation
ReLU (Rectified Linear Unit)	$f(x) = \alpha * (x - threshold) \quad x < threshold$ $f(x) = x \quad threshold \leq x \leq max_value$ $f(x) = max_value \quad x \geq max_value$	
ELU (Exponential linear unit)	$f(x) = \alpha * (\exp(x) - 1) \quad x < 0$ $f(x) = x \quad x \geq 0$	
TanH (Hyperbolic tangent)	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	
Sigmoid	$f(x) = \frac{1}{1 + e^{-x}}$	
Hard_sigmoid	$f(x) = 0 \quad x < -2.5$ $f(x) = 0.2 * x + 0.5 \quad -2.5 \leq x \leq 2.5$ $f(x) = 1 \quad x \geq 2.5$	
SoftPlus	$f(x) = \log(\exp(x) + 1)$	
Linear	$f(x) = x$	

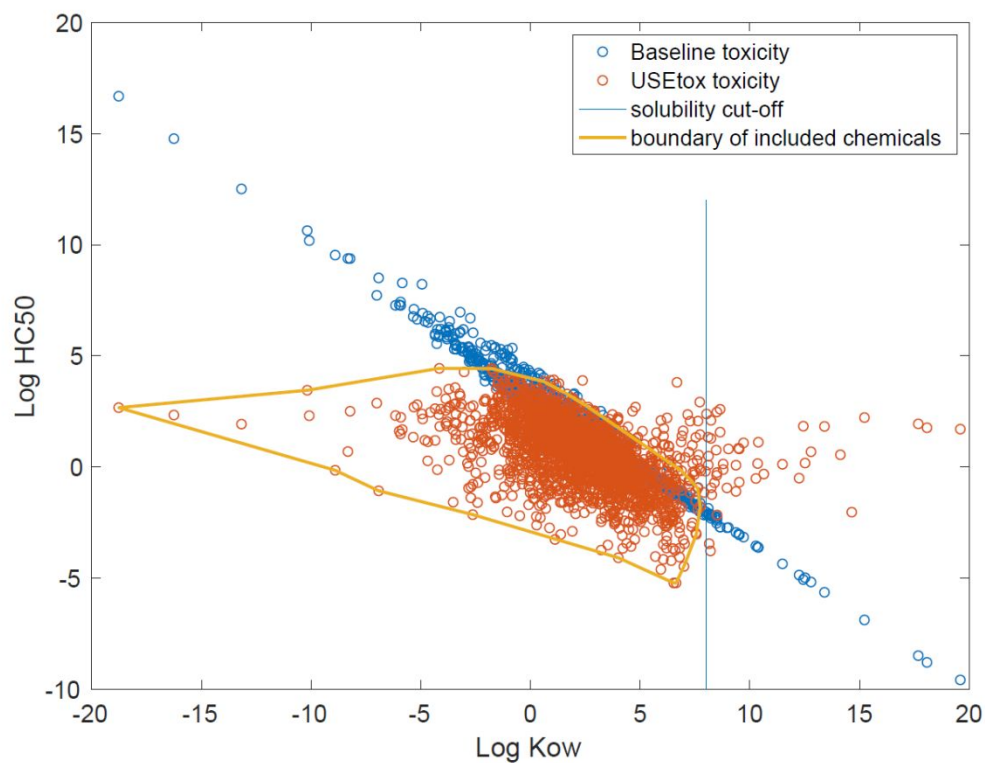


Figure S1. Data filtering result

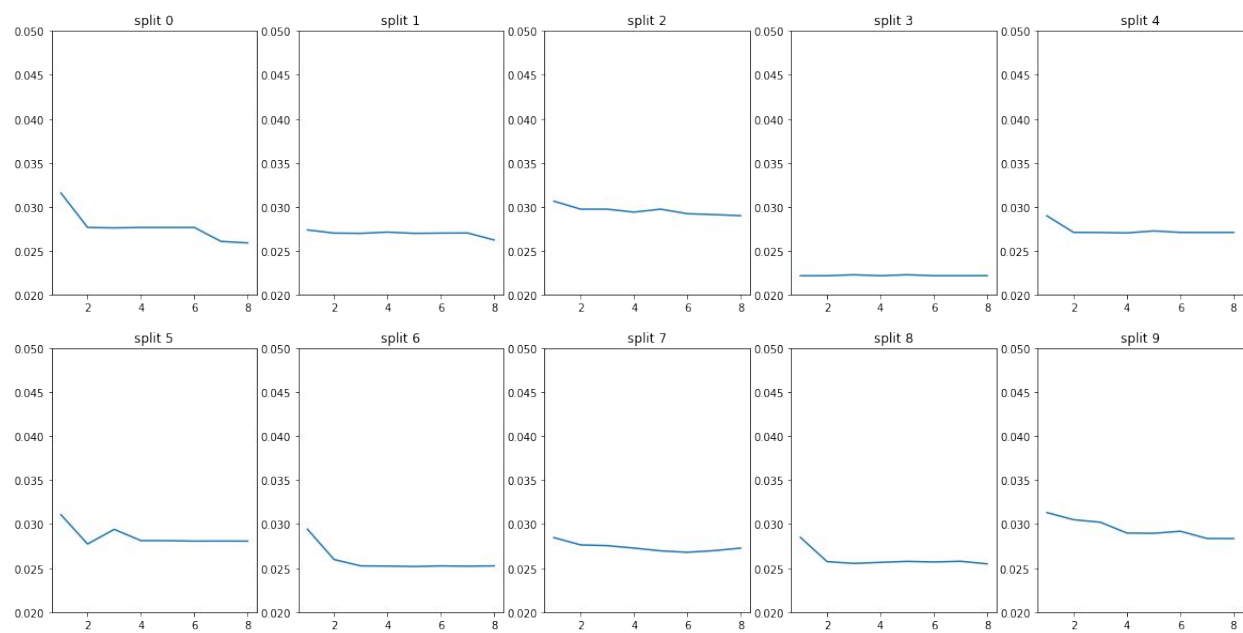


Figure S2. Fitness (i.e., validation MSE) of the best model along eight generations.