Supplementary Information

Benchmarking Deep Learning Models for Predicting Anticancer Drug Potency (IC_{50}): Insights for the Medicinal Chemist

Udbhas Garai[1†], Aditya S. Pal[2†], Koyel Ghosh[2], Deepak B. Salunke[3] and Utpal Garain[2,4*]

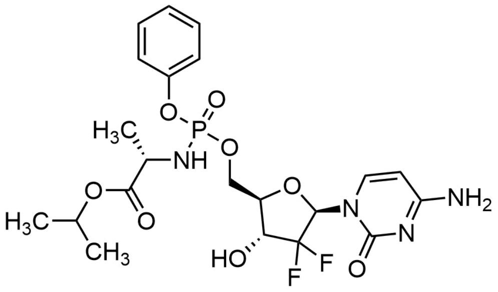
[1]Department of Chemical Sciences, Indian Institute of Science Education and Research Mohali, Punjab 140306, India
[2]Computer Vision & Pattern Recognition [CVPR] Unit, Indian Statistical Institute, Kolkata, West Bengal 700108, India
[3]Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research, S.A.S. Nagar, Punjab 160062, India
[4]Centre for Artificial Intelligence and Machine Learning (CAIML), Indian Statistical Institute, Kolkata, West Bengal 700108, India

*Correspondence: utpal@isical.ac.in

[†]These authors contributed equally to this work.

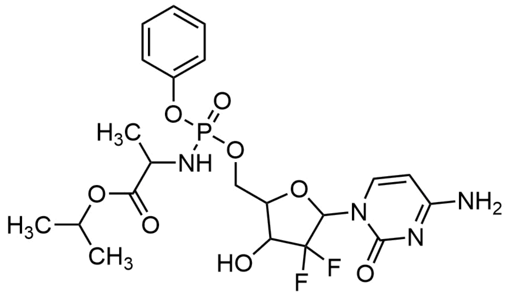
Contents		

a



NC1=NC(N(C=C1)[C@H]2C(F)(F)[C@H](O)[C@@H](COP
(OC3=CC=CC=C3)(N[C@@H](C)C(OC(C)C)=O)=O)O2)=O

b



NC1=NC(N(C=C1)C2C(F)(F)C(O)C(COP(OC3=CC=C
C=C3)(NC(C)C(OC(C)C)=O)=O)O2)=O

Supplementary Fig. 1: The two SMILES representations. (a) Isomeric SMILES representation that encodes that stereochemical (chiral) information of the molecule. “@” lists neighboring atoms anticlockwise; “@@” lists them clockwise. (b) Canonical or Generic SMILES represention that does not have any stereochemical information. No “@” or “@@” symbols used.

Supplementary Table 1: Prediction performance of four deep learning models on unseen anticancer compounds from literature using isomeric and canonical (or generic) SMILES representations.

Model	Isomeric SMILES				Canonical/Generic SMILES			
	r	$MAPE(IC_{50})$ (%)	EVAPA (%)	within $3\sigma_{IC_{50}}$ (%)	r	$MAPE(IC_{50})$ (%)	EVAPA (%)	within $3\sigma_{IC_{50}}$ (%)
DeepCDR	0.6381	2268.44	37.50	12.50	0.6291	2346.34	40.63	15.63
DrugCell	0.4829	2507.52	37.50	18.75	0.5246	1167.97	56.25	9.38
PaccMann	−0.4650	6836.81	31.25	6.25	−0.3540	3445.20	34.38	6.25
Precily	0.6384	2920.53	40.63	12.50	0.5311	4205.95	34.38	6.25

Proof of $\sigma(\Delta x) = \sqrt{2} \times \sigma(x)$ used in EVAPA derivation

Let x_1 and x_2 be 2 independent variables. Using the property of variance for sums of independent variables, we can write:

$$\mathrm{Var}(x_1 - x_2) \;=\; \mathrm{Var}(x_1) + \mathrm{Var}(x_2) - 2\,\mathrm{Cov}(x_1, x_2)$$

Since x_1 and x_2 are independent, $\mathrm{Cov}(x_1, x_2) = 0$. Thus,

$$\begin{aligned}
& \mathrm{Var}(x_1 - x_2) = \mathrm{Var}(x_1) + \mathrm{Var}(x_2) - 2\,\mathrm{Cov}(x_1, x_2) \\
\implies\; & \mathrm{Var}(x_1 - x_2) = \mathrm{Var}(x_1) + \mathrm{Var}(x_2) \qquad (\because \mathrm{Cov}(x_1, x_2) = 0) \\
\implies\; & \mathrm{Var}(\Delta x) = \mathrm{Var}(x_1) + \mathrm{Var}(x_2) \\
\implies\; & \sigma^2(\Delta x) = \sigma^2(x_1) + \sigma^2(x_2) \\
\implies\; & \sigma^2(\Delta x) = 2\,\sigma^2(x) \\
\implies\; & \sigma(\Delta x) = \sqrt{2}\,\sigma(x)
\end{aligned}$$