Table S15(a): Results of statistical significance tests between ImageMol and Image-based methods (Chemception, ADMET-CNN, QSAR-CNN) on five CYP450 isoforms training sets (PubChem Data Set I) with 5-fold cross-validation, where numbers represent the p-values (one-sided significance level) of the McNemar's test. The numbers in green background indicate statistically different models, using a significance threshold of 0.05. 0 indicates statistical significance less than E-100.

PubChem Data Set I	Chemception	ADMET-CNN	QSAR-CNN
CYP1A2	8.31E-46	0	0
CYP2C9	9.32E-39	7.49E-95	1.22E-95
CYP2C19	3.07E-42	0	0
CYP2D6	6.48E-42	2.36E-38	2.36E-38
CYP3A4	9.89E-18	2.47E-18	1.94E-161

—— continue ——

Table S15(b): Results of statistical significance tests between ImageMol and Image-based methods (Chemception, ADMET-CNN, QSAR-CNN) on five CYP450 isoforms validation sets (PubChem Data Set II), where numbers represent the p-values (one-sided significance level) of the McNemar's test. The numbers in green background indicate statistically different models, using a significance threshold of 0.05. 0 indicates statistical significance less than E-100.

PubChem Data Set II	Chemception	ADMET-CNN	QSAR-CNN
CYP1A2	1.23E-30	0	2.56E-28
CYP2C9	4.43E-36	3.26E-20	1.78E-31
CYP2C19	3.07E-20	1.94E-23	4.62E-15
CYP2D6	2.41E-21	4.94E-20	2.21E-25
CYP3A4	5.27E-16	0	2.27E-36