Supplementary Materials.

Table S1. PubChem bioassay activity class values and descriptions used to classify compounds as "active" or "inactive."

Activity Class Value	Activity Class Description	Categorization
1.1	Complete response, efficacy >80%, R2 ≥ 0.9	Active
1.2	Partial response, efficacy ≤ 80%, R2 ≥ 0.9	Active
2.1	Incomplete curve, efficacy >80%, R2 > 0.9	Active
4.0	Inactive	Inactive

Table S2. resulting number of compounds from individual dataset after preprocessing.

CYP450 isozymes	Data sets	Number of inhibitors	Number of non-inhibitors
1A2	PubChem AID 1851	4396	6642
	PubChem AID 410	1723	238
	ChEMBL database	579	1149
	Combined data	**6698**	**8029**
2C9	PubChem AID 1851	2951	7999
	PubChem AID 883	61	476
	ChEMBL database	1280	1604
	Combined data	**4292**	**10079**
2C19	PubChem AID 1851	4949	6801
	PubChem AID 899	141	422
	ChEMBL database	611	1066
	Combined data	**5701**	**8289**
2D6	PubChem AID 1851	1552	10502
	PubChem AID 891	61	542
	ChEMBL database	1390	1902
	Combined data	**3003**	**12946**
3A	PubChem AID 1851	3443	7196
	PubChem AID 884	586	1952
	ChEMBL database	2576	2274
	Combined data	**6605**	**11422**

Table S3. Optimized parameters for machine learning models

Algorithm	Parameters

Random Forest	max_depth = None,
	max_features = sqrt, min_samples_leaf = 2, min_samples_split = 5,
	n_estimators = 100

Support Vector Machine	C= 10.0,
	Gamma= 0.01,
	Kernel= rbf,
	probability: True

LightGBM	boosting_type = dart,
	learning_rate = 0.1,
	'max_depth = -1,
	n_estimator = 300,
	num_leaves = 31

K-Nearest Neighbors	n_neighbors = 10, P (Minkowski metric) = 1, Weights = distance

Table S4. Results of 10-Fold cross-validation for Five CYP450 isoforms using each model

Isozymes	**Metrics** / **Models**	**Accuracy**	**AUC**	**TPR**	**TNR**	**F-measure**	**MCC**
1A2	LGBM	0.84	0.92	0.80	0.88	0.82	0.68
	SVC	0.84	0.91	0.80	0.88	0.82	0.68
	RF	0.83	0.91	0.76	0.88	0.80	0.65
	XGBC	0.82	0.90	0.76	0.87	0.79	0.63
	KNN	0.82	0.90	0.81	0.82	0.80	0.64
	GNB	0.72	0.76	0.76	0.69	0.71	0.45
2C9	SVC	0.86	0.92	0.72	0.92	0.75	0.66
	LGBM	0.86	0.92	0.72	0.92	0.775	0.66
	RF	0.84	0.91	0.63	0.93	0.71	0.61
	KNN	0.84	0.90	0.61	0.93	0.69	0.59
	XGBC	0.82	0.88	0.65	0.90	0.69	0.57
	GNB	0.67	0.75	0.85	0.59	0.61	0.41
2D6	SVC	0.91	0.92	0.65	0.97	0.73	0.68
	LGBM	0.90	0.92	0.62	0.97	0.71	0.67
	KNN	0.89	0.90	0.59	0.96	0.68	0.63
	RF	0.89	0.92	0.51	0.98	0.64	0.62
	XGBC	0.88	0.89	0.50	0.98	0.62	0.58
	GNB	0.63	0.72	0.83	0.58	0.46	0.32
3A4	SVC	0.87	0.94	0.79	0.91	0.81	0.72
	LGBM	0.86	0.93	0.79	0.91	0.81	0.70
	RF	0.85	0.93	0.73	0.92	0.78	0.68
	XGBC	0.83	0.90	0.73	0.88	0.76	0.62
	KNN	0.84	0.91	0.71	0.92	0.76	0.65
	GNB	0.73	0.79	0.83	0.67	0.69	0.48
2C19	LGBM	0.84	0.91	0.80	0.86	0.80	0.66
	SVC	0.84	0.91	0.79	0.88	0.80	0.67
	RF	0.83	0.90	0.77	0.87	0.78	0.64
	KNN	0.81	0.89	0.76	0.85	0.77	0.61
	XGBC	0.80	0.88	0.78	0.82	0.76	0.60
	GNB	0.72	0.76	0.82	0.64	0.70	0.46

Table S5. Results of each model on the Test sets for the 5 CYP450 isoforms.

Isozymes	Metrics Models	Accuracy	AUC	TPR	TNR	F-measure	MCC
1A2	LGBM	0.85	0.92	0.80	0.88	0.83	0.69
	SVM	0.85	0.91	0.81	0.88	0.83	0.69
	ANN	0.84	0.84	0.79	0.88	0.84	0.68
	RF	0.83	0.91	0.77	0.88	0.81	0.66
	XGB	0.82	0.90	0.77	0.87	0.80	0.64
	KNN	0.81	0.89	0.80	0.82	0.80	0.62
	GNB	0.72	0.75	0.76	0.68	0.71	0.44
2C9	SVM	0.86	0.91	0.74	0.91	0.76	0.66
	LGBM	0.86	0.91	0.73	0.91	0.75	0.65
	ANN	0.85	0.81	0.72	0.90	0.85	0.63
	RF	0.83	0.90	0.63	0.92	0.69	0.58
	KNN	0.83	0.88	0.61	0.92	0.67	0.56
	XGB	0.82	0.88	0.64	0.89	0.67	0.55
	GNB	0.66	0.74	0.85	0.58	0.60	0.40
2D6	SVM	0.91	0.92	0.64	0.97	0.72	0.67
	LGBM	0.91	0.92	0.62	0.97	0.71	0.67
	KNN	0.90	0.91	0.60	0.97	0.70	0.66
	RF	0.90	0.92	0.52	0.99	0.66	0.64
	ANN	0.90	0.79	0.61	0.96	0.89	0.64
	XGB	0.88	0.89	0.46	0.98	0.59	0.56
	GNB	0.63	0.72	0.81	0.59	0.45	0.31
3A4	SVM	0.86	0.93	0.78	0.91	0.81	0.70
	ANN	0.86	0.84	0.80	0.89	0.86	0.69
	LGBM	0.85	0.93	0.78	0.89	0.79	0.68
	RF	0.85	0.92	0.72	0.92	0.78	0.66
	XGB	0.82	0.89	0.72	0.89	0.74	0.60
	KNN	0.83	0.91	0.69	0.91	0.75	0.63
	GNB	0.73	0.78	0.83	0.67	0.69	0.48
2C19	LGBM	0.84	0.91	0.80	0.87	0.81	0.68
	SVM	0.84	0.91	0.79	0.88	0.80	0.67
	RF	0.83	0.90	0.76	0.88	0.79	0.65
	ANN	0.82	0.81	0.75	0.87	0.82	0.63
	KNN	0.82	0.89	0.75	0.86	0.77	0.62
	XGB	0.81	0.88	0.80	0.82	0.77	0.61
	GNB	0.70	0.74	0.81	0.63	0.69	0.43

Table S6. Performance comparison between proposed work using SVM (M) and previous studies including CYPlebrity [9], Cheng et al. [10] (SVM + C4.5DT), and WhichCyp [11].

Isozymes	**Metrics** **Papers**	**Accuracy**	**AUC**	**TPR**	**TNR**	**F-measure**	**MCC**
1A2	M	0.85	0.91	0.81	0.88	0.83	0.69
	[3]	0.82	0.90	0.81	0.83	0.82	0.64
	[10]	0.72	0.81	0.77	0.65	-[1]	0.41
	[9]	0.88	0.95	0.87	0.88	-	-
2C9	M	0.86	0.91	0.74	0.91	0.76	0.66
	[3]	0.85	0.92	0.70	0.93	0.76	0.65
	[10]	0.86	0.85	0.56	0.96	-	0.60
	[9]	0.84	0.90	0.84	0.83	-	-
2D6	M	0.91	0.92	0.64	0.97	0.72	0.67
	[3]	0.90	0.92	0.69	0.96	0.75	0.70
	[10]	0.88	0.87	0.58	0.94	-	0.57
	[9]	0.84	0.88	0.75	0.86	-	-
3A4	[M]	0.86	0.93	0.78	0.91	0.81	0.70
	[3]	0.85	0.92	0.74	0.92	0.80	0.68
	[10]	0.75	0.78	0.46	0.87	-	0.35
	[9]	0.84	0.92	0.84	0.84	-	-
2C19	M	0.84	0.91	0.80	0.87	0.80	0.67
	[3]	0.81	0.89	0.82	0.81	0.79	0.62
	[10]	0.81	0.84	0.51	0.91	-	0.47
	[9]	0.85	0.91	0.86	0.84	-	-

[1] - : Means the value was not provided