

Table 7. Performance of individual methods and the LLNA for predicting human skin sensitization hazard compared with machine learning approaches

Method	Data set ^a	Accuracy (%)	Sensitivity (%)	Specificity (%)
Machine learning models ^b	Training	93–94	92–94	95
	Test	92	87–93	89–100
	All	93–94	92–94	94–96
h-CLAT	Training	82	88	67
	Test	79	87	67
	All	81	88	67
DPRA	Training	88	88	86
	Test	71	73	67
	All	83	85	80
KeratinoSens	Training	78	82	67
	Test	63	60	67
	All	74	77	67
Toolbox	Training	81	82	76
	Test	71	73	67
	All	78	80	73
LLNA	Training	83	90	67
	Test	88	100	67
	All	84	92	67
Test battery 1 (≥ 1 positive = positive)	Training	79	98	33
	Test	75	100	33
	All	78	99	33

Method	Data set	Accuracy (%)	Sensitivity (%)	Specificity (%)
Test battery 2 (≥ 2 positives = positive)	Training	89	96	71
	Test	75	87	56
	All	85	94	67

DPRA, direct peptide reactivity assay; h-CLAT, human cell line activation test; LLNA, murine local lymph node assay; Toolbox, read-across using QSAR Toolbox.

a Test set contains 15 sensitizers and nine non-sensitizers. The training set contains 51 sensitizers and 21 non-sensitizers. "All" is the entire data set of 96 substances: 66 sensitizers and 30 non-sensitizers.

b Models with the highest performance from Table 4: support vector machine and logistic regression models with variable groups A, I and K.