Supplementary Materials

Model Comparison for Learning Curve Fitting

While the main text presents results based only on the LOESS model, additional curve-fitting analyses were performed to test how sensitive the learning-curve pattern was to different functional forms (Figure S1; Table S1). We compared four alternative models—logarithmic, power-law, Michaelis–Menten, and exponential saturation—to estimate the training sample size required to reach 90 % and 95 % of the observed maximum performance (61% for overall accuracy and 56% for macro F1 score).

Although all models showed the general trend of rapid improvement followed by saturation, the thresholds varied across models. In general, the logarithmic and power-law fits reached their plateaus more slowly, requiring larger training areas, whereas the Michaelis–Menten and exponential saturation models showed faster early gains. Notably, Michaelis–Menten and exponential saturation models failed to reach the 95% threshold within the observed data range, as their fitted curves plateaued below this target value—these cases are indicated as "NA" in Table S1. This limitation highlights the importance of model selection when extrapolating learning curves beyond the observed data range.

Among them, the LOESS fit provided the best agreement with the observed data, showing the highest R^2 values. Figure S1 illustrates the fitted curves and their 90 %/95 % thresholds, and Table S1 summarizes the corresponding R^2, maximum predicted values, and values of 90 %/95 % thresholds.

Table S1: Model fitting performance and estimated training sample size thresholds for four curve types. The table presents the coefficient of determination (R^2), the maximum predicted performance, and the training sample size (%) required to reach 90 % and 95 % of the maximum value.

Metric	Model	R_squared	Training sample size at 90pct of Max	Training sample size at 95pct of Max
Overall Accuracy	LOESS	0.89	2.67	5.89
	logarithmic	0.88	5.59	9.99
	power law	0.83	6.33	10.43
	Michaelis menten	0.81	3.11	NA
	Exponential saturation	0.64	1.06	NA
Macro F1 Score	LOESS	0.91	2.37	4.42
	logarithmic	0.89	5.01	7.94
	power law	0.82	5.74	8.67
	Michaelis menten	0.89	2.37	8.08
	Exponential saturation	0.80	1.2	NA

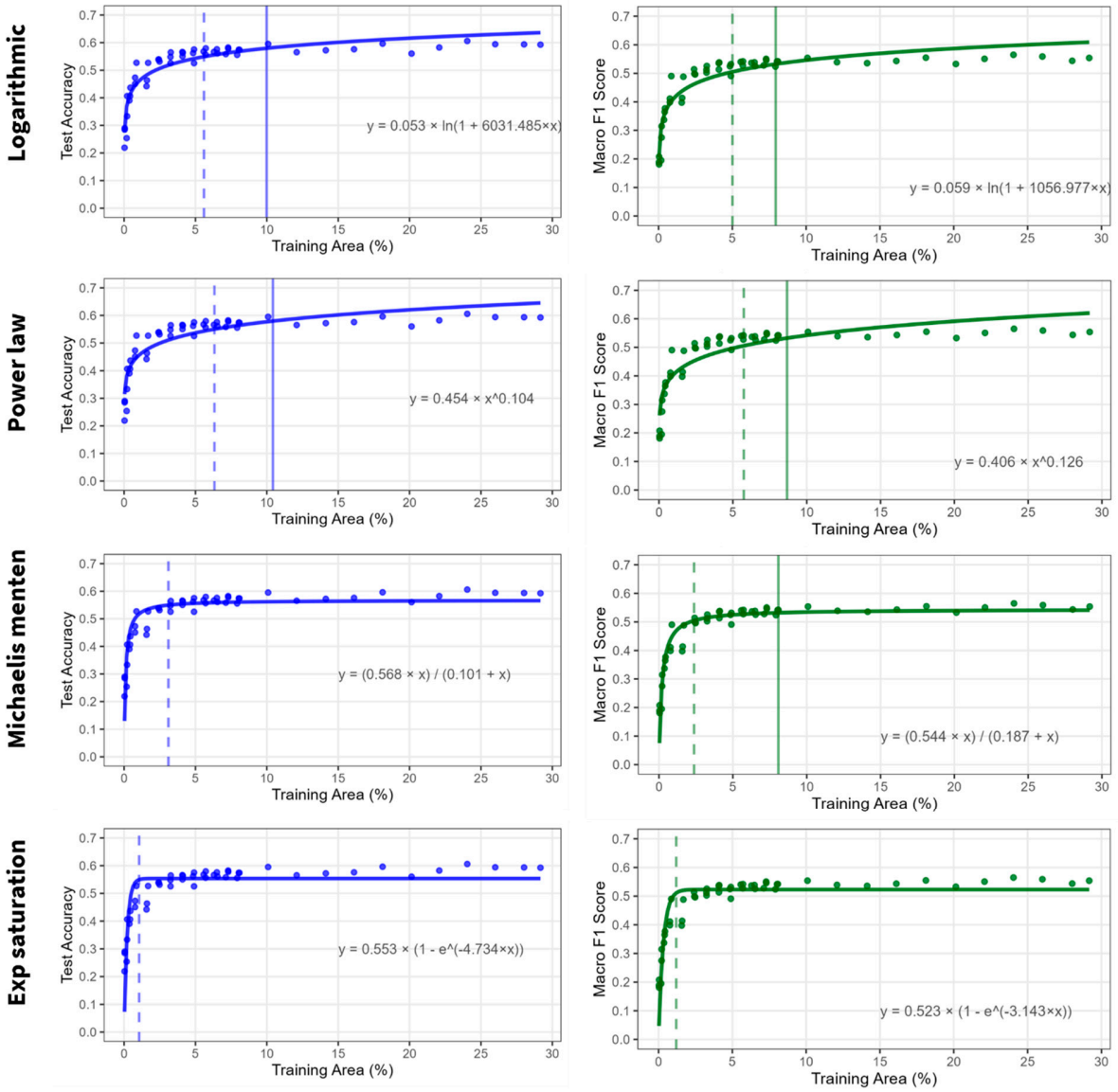


Figure S1. Relationship between training sample size (training area %) and model performance. Each curve represents the fitted function for four model types. Vertical dashed lines indicate the training area at which the model achieves 90 % of its maximum score, and vertical solid lines indicate the point corresponding to 95 % of the maximum score. Left panels show Test Overall Accuracy, and right panels show Macro F1 Score.