Supplementary methods and numerical results

Credibility assessment of machine learning models for slotted cold-formed steel beams

This supplement identifies the numerical evidence supporting the manuscript. All tables refer to the completed full-profile run with 14 models, two targets and five partitions. It accompanies the Supplementary_Data directory and does not combine results with the reduced execution-check profile.

S1 Reproducibility

The source and data fingerprints, software versions and complete configuration are recorded in run_manifest.json. stage_status.json records completion of all twelve stages. Reference prediction_records.csv contains 22,596 unique target–model–scenario–case predictions; conformal summaries describe 1,355,760 interval-case evaluations over thirty allocations at two nominal levels. Source FE table matching agrees within 0.005 kN m for all 432 cases of each target.

S2 Experimental partitions

split_assignments.csv contains each reference training/test allocation. conformal_split_assignments.csv records proper-training, calibration and test IDs for each allocation; training_only_cv_fold_ids.csv records the training-only fit/validation folds. These files permit direct checking that test responses are excluded from fitting, calibration and parameter selection.

S3 Model configuration and optimization

fit_audit.csv reports initial and fitted reference-model parameters and warnings. training_only_cv_candidates.csv and training_only_cv_selected.csv distinguish candidate configurations, CV scores and selected settings. Elastic Net convergence warnings occur in some candidate folds; all ten selected Elastic Net configurations were refitted separately and converged. GPR warnings identify kernel-bound contacts in candidate fits. They are reported rather than suppressed and the compact grid is not asserted to establish an algorithm-wide optimum.

S4 Accuracy and overprediction

performance_summary.csv gives MAE, RMSE, MAPE, R squared, signed error, overprediction frequency, maximum relative overprediction, severe-overprediction counts and tolerance attainment for every method and scenario. largest_relative_overpredictions.csv identifies the corresponding worst overpredicted cases. mae_inflation_verified.csv uses the same unrounded reference source.

S5 Training size and matched transfer

size_control_summary.csv provides thirty allocations at each of 216, 288 and 345 ID training cases. The full-size control reuses the identical ID fit, while smaller subsets retain the original ordering. matched_perforation_baseline.csv reports direct response transfer between 216 matched pairs in both directions. target_distributions.csv gives scenario training/test response composition.

S6 Prediction intervals

conformal_summary.csv contains nominal 90% and 95% coverage counts, Wilson reference bounds, widths, ranks and accepted-case outcomes for every model, scenario and allocation. conformal_repeated_split_summary.csv gives mean, SD and range across allocations. Each calibration subset is within the corresponding training domain and the predictor is fitted on proper training only. Endpoint ties count as covered; an unattainable finite-sample rank yields an unbounded interval. Shared test cases across allocations are not pooled as independent experiments.

S7 Applicability and retention

domain_case_checks.csv and domain_threshold_sensitivity_cases.csv record marginal, covariance, schema and nearest-combination diagnostics. domain_error_outcomes.csv supplies accepted and rejected errors at the four prespecified quantiles. Support thresholds use training inputs only. Zero retained OOD queries means accepted-case accuracy and interval coverage are undefined, not zero-error predictions.

S8 Estimator and feature sensitivity

estimator_seed_summary.csv and estimator_seed_stability.csv separate stochastic repetitions from deterministic reference fits. feature_sensitivity_summary.csv gives full, dropped-radius, dropped-thickness and dropped-constant refits. SHAP provenance files specify the fitted training domain, background/evaluation IDs, explanation setting and additivity checks. The fixed-ID-model population diagnostic is descriptive and records any overlap of its evaluated population with ID training.

Table S1 Fitted GPR kernels for the reference experiments

Mode	Scenario	Fitted kernel
LB	ID	1.59**2 * RBF(length_scale=5.88) + WhiteKernel(noise_level=3.71e-05)
DB	ID	2.57**2 * RBF(length_scale=7.86) + WhiteKernel(noise_level=0.000484)
LB	Geometry	1.6**2 * RBF(length_scale=6.64) + WhiteKernel(noise_level=1.67e-05)
DB	Geometry	2**2 * RBF(length_scale=7.75) + WhiteKernel(noise_level=0.000474)
LB	Slots 75 to 60	1.61**2 * RBF(length_scale=5.54) + WhiteKernel(noise_level=4.63e-05)
DB	Slots 75 to 60	2.62**2 * RBF(length_scale=7.57) + WhiteKernel(noise_level=0.000675)
LB	Material	1.97**2 * RBF(length_scale=7.75) + WhiteKernel(noise_level=4.29e-05)
DB	Material	2.77**2 * RBF(length_scale=9.76) + WhiteKernel(noise_level=0.000517)
LB	Slots 60 to 75	1.7**2 * RBF(length_scale=5.68) + WhiteKernel(noise_level=3.44e-05)
DB	Slots 60 to 75	2.64**2 * RBF(length_scale=7.65) + WhiteKernel(noise_level=0.000343)

Table S2 Training only validation sensitivity MAEs

Mode	Model	ID	Geometry	Slots 75→60	Material	Slots 60→75
LB	Ridge	2.774	4.908	2.714	4.063	2.729
LB	Lasso	2.747	4.907	2.707	4.116	2.715
LB	Elastic Net	2.776	4.907	2.713	4.057	2.727
LB	Decision Tree	0.117	14.718	0.163	5.868	0.163
LB	kNN	2.808	6.744	1.956	5.872	1.922
LB	SVR	0.155	4.425	9.035	8.179	8.990
LB	GPR	0.083	2.593	8.898	2.123	8.851
LB	Random Forest	0.096	9.510	0.156	5.869	0.172
LB	GBRT	0.070	14.657	0.163	5.869	0.161
LB	HistGBR	0.342	12.906	0.283	5.873	0.279
LB	AdaBoost	2.093	14.052	1.812	6.710	1.696
LB	XGBoost	0.061	13.813	0.163	5.868	0.163

Mode	Model	ID	Geometry	Slots 75→60	Material	Slots 60→75
LB	LightGBM	0.081	13.586	0.169	5.868	0.167
LB	CatBoost	0.067	5.918	0.163	5.868	0.163
DB	Ridge	2.247	4.108	2.234	3.074	2.217
DB	Lasso	2.245	4.108	2.232	3.085	2.240
DB	Elastic Net	2.248	4.108	2.234	3.058	2.217
DB	Decision Tree	0.201	12.354	0.192	4.275	0.192
DB	kNN	2.229	5.939	1.598	4.294	1.586
DB	SVR	0.143	3.862	6.871	6.425	7.383
DB	GPR	0.200	2.106	6.463	1.300	6.502
DB	Random Forest	0.133	8.015	0.190	4.269	0.201
DB	GBRT	0.100	12.200	0.192	4.276	0.193
DB	HistGBR	0.361	10.730	0.264	4.284	0.269
DB	AdaBoost	1.607	8.065	1.389	4.758	1.541
DB	XGBoost	0.105	8.782	0.192	4.275	0.192
DB	LightGBM	0.174	11.792	0.202	4.275	0.216
DB	CatBoost	0.109	5.191	0.192	4.275	0.197