Appendix 2. Comparison of Results

Research Paper	Accuracy	F1 Score	R^2	MAE	Notes
Machine learning and soft voting ensemble classification for earthquake-induced damage to bridges [1]	0.79 (ANN), 0.79 (SVM), 0.76 (RF), 0.80 (Ensemble)	-	-	-	Homogeneous data (RCC bridges) yielded better accuracies for models like RF (0.83) and CatBoost (0.81) compared to the mixed database.
Machine learning-based collapse prediction for post-earthquake damaged RC columns under subsequent earthquakes [2]	0.82	0.76	-	-	SHAP analysis identified axial compression ratio and other structural performance factors as critical for collapse prediction.
VHXLA: A post-earthquake damage prediction method for high-speed railway track-bridge system using VMD and hybrid neural network [3]	0.9996 (combination of models)	0.945 (combination of models)	-	-	The VHXLA model significantly outperforms other models in prediction accuracy and F1 scores for sliding layers and fixed bearings, demonstrating superior predictive ability for seismic damage. The inclusion of VMD decomposition, HT transformation, and the attention mechanism in the VHXLA model enhances its performance compared to simpler models like LSTM or Xception.
Seismic damage state predictions of reinforced concrete structures using stacked long short-term memory neural networks [4]	0.95 (4-story, 90 stacks), 0.93 (8-story, 90 stacks), 0.96 (12-story, 90 and 150 stacks)	-	-	-	The proposed LSTM model effectively generalized across different ductile frame heights (4-, 8-, and 12-story frames), providing reliable post-earthquake damage predictions.
Seismic response prediction and fragility assessment of high-speed railway bridges using machine learning technology [5]	-	-	>0.847 (y_2, y_3)	0.61 (y_1), 0.53 (y_4)	RF, XGBoost, and Light GBM consistently outperform Lasso, ANN, and SVR in both regression performance and error metrics.
Post-earthquake seismic capacity estimation of reinforced concrete bridge piers using machine learning techniques [6]	-	-	0.736 (DT), 0.277 (KNN), 0.166 (Lasso regression), 0.153 (Linear regression), 0.162 (Ridge regression), 0.146 (SVM regression), 0.179 (ANN), 0.121 (RF)	3.284* (DT), 4.896* (KNN), 5.873* (Lasso), 5.201* (ANN), 4.946* (SVM regression), 6.002* (Linear regression), 5.935* (Ridge regression), 6.464* (RF); * - SMAPE (Symmetric Mean Absolute Percentage Error) is provided as a related error metric.	The ensemble model (RF) had lower R^2 and higher SMAPE, underperforming compared to DT. Simpler models (LR, RR, Lasso) also showed significantly lower predictive accuracy and precision compared to complex models like DT, KNN, and ANN.
Deep autoencoder architecture for bridge damage assessment using responses from several vehicles [7]	Effective damage detection even with random traffic, though severity quantification is slightly impacted by random traffic contributions.	-	-	distributions shift significantly as damage severity increases	Minimal MAE for undamaged states. Sensitivity to stiffness reduction (5-30%). Performs well under varying conditions.
Automated location of steel truss bridge damage using machine learning and raw strain sensor data [8]	0.73 (CNN3), 0.75 (CNN4), 0.56 (1NN-DTW)	-	-	-	CNNs are clearly superior to 1NN-DTW models for both damage location and severity assessment. Damage Severity Assessment is more challenging, with all models achieving lower accuracies compared to damage location.
Seismic damage identification of high arch dams based on an unsupervised deep learning approach [9]	0.79316 (optimal)	0.85749 (dropout ratio = 0.35)			
Application of machine learning in seismic fragility assessment of bridges with SMA-restrained rocking columns [10]	-	-	0.78 - 0.97 (ANN), 0.76 - 0.96 (SVR), 0.72 - 0.95 (Ridge), 0.76 - 0.96 (AdaBoost), 0.78 - 0.95 (RF)	< 0.12* (ANN, 0.31 for Max. ED link damage factor), < 0.16* (SVR, 0.35 for Max. ED link damage factor), < 0.18* (Ridge, 0.41 for Max. ED link damage factor), < 0.21* (AdaBoost, 0.27 for Max. ED link damage factor), < 0.18* (RF, 0.32 for Max. ED link damage factor), * - Mean Squared Error (MSE)	Support vector regression (SVR), Ridge regression, AdaBoost, and Random forest also performed well, but the R^2 values were generally slightly lower than those of the neural networks.