
Supplementary Information

Stacking Ensemble Learning with Regression Models for Predicting Damage from Terrorist Attacks

---

1. Hyperparameter Tuning for Classification and Regression Models

1.1 Classification Models

To select base classifiers, seven machine learning (ML) models, including k-Nearest Neighbor (KNN), Logistic Regression (LR), Support Vector Machine (SVM), Multilayer Perceptron (MLP), Random Forest (RF), Gradient Boosting Machine (GBM), and Categorical Boosting (CatBoost), with default hyperparameter values were compared their performance in the classification of fatal, injury, and property damage attacks. The top three best classifiers were selected to create meta-models for identifying fatal attacks, injury attacks, and property damage attacks. As a result, the top three models selected are CatBoost, GBM, and RF.

After that, grid search was used to find suitable hyperparameter values of those three models. The lists of the considered hyperparameter values in all classification cases (i.e., fatal, injury, and property damage attacks) are shown in Table S1.

Table S1: *The considered hyperparameter values of the classifiers for grid search.*

Classifiers	Hyperparameters	Considered Values
RF	- The number of decision trees (*n_estimators*)	100, 200, 300, and 500 $^{a, b, c}$
GBM	- The number of decision trees (*n_estimators*)	100, 500, 700 $^{a, b}$, and 1000 ^{c}
	- Learning rate (*learning_rate*)	0.05, 0.10 $^{a, b, c}$, 0.15, and 0.20
CatBoost	- The number of decision trees (*iterations*)	500, 1000, 1500, 2000 $^{a, c}$, and 2500 ^{b}
	- Learning rate (*learning_rate*)	0.04, 0.05, 0.07, 0.10 $^{a, b}$, 0.15 ^{c}, and 0.20

The subscripts a, b, and c indicate the appropriate hyperparameter values that contribute the best performance values in the classification of fatal attacks, injury attacks, and property damage attacks, respectively.

1.2 Regression Models

Five regression models, including Ridge Regression (RR), Multilayer Perceptron (MLP), Random Forest (RF), Gradient Boosting Machine (GBM), and Categorical Boosting (CatBoost), were compared their predicting performance to select the best one. In grid search, the considered hyperparameter values of all regression models are shown in Table S2.

Table S2: *The considered hyperparameter values of the regression models for grid search.*

Regression Models	**Hyperparameters**	**Considered Values**
RR	- The constant controlling prediction errors (*alpha*)	0.5, 1, 10, 100, 500, 1000, 3000, 5000 ^{d}, 7500, and 10000 ^{e}
MLP	- The number of hidden nodes (*hidden_layer_sizes*)	20, 30, 50 $^{d, e}$, 100, and 200
RF	- The number of decision trees (*n_estimators*)	20, 50, 100, 200, and 300 $^{d, e}$
GBM	- The number of decision trees (*n_estimators*)	100, 500, 700, and 1000 $^{d, e}$
	- Learning rate (*learning_rate*)	0.005, 0.010 ^{d}, 0.050 ^{e}, and 0.100
CatBoost	- The number of decision trees (*iterations*)	500, 1000, 1500, and 2000 $^{d, e}$
	- Learning rate (*learning_rate*)	0.01, 0.05, 0.10, and 0.15 $^{d, e}$

The subscripts d and e indicate the appropriate hyperparameter values that contribute the best performance values in estimation of fatalities and injuries, respectively.

2. Performance Comparison of Different Meta-Models

In this experiment, different meta-models or stacking ensemble models combining different base classifiers were compared their performance in the classification of fatal, injury, and property damage attacks to justify that selecting the top three best models is the most suitable approach. Three different methods to combine base classifiers are 1) combining the top three best classifiers, 2) combining the three diverse performing classifiers (the best one, the middle one, and the worst one), and 3) combining the bottom three worst classifiers. F1 scores were used to rank seven different classifiers in the classification of fatal, injury, and property damage attacks, and the average rank of each classifier over all classification cases can be computed as shown in Table S3. The average F1 scores of three meta-models in the classification of fatal, injury, and property damage attacks are shown in Table S4.

Table S3: *Average F1 scores and ranks of seven classification models.*

Classification Models	Classification Cases			Average Rank
	Fatal Attacks	Injury Attacks	Property Damage Attacks	
KNN	0.791 (6)	0.678 (5)	0.771 (7)	6.0 (7)
LR	0.792 (5)	0.653 (6)	0.781 (5)	5.3 (5)
SVM	0.791 (6)	0.646 (7)	0.782 (4)	5.7 (6)
MLP	0.798 (4)	0.692 (4)	0.778 (6)	4.7 (4)
RF	0.808 (3)	0.695 (3)	0.802 (3)	3.0 (3)
GBM	0.817 (2)	0.703 (2)	0.809 (2)	2.0 (2)
CatBoost	0.821 (1)	0.710 (1)	0.813 (1)	1.0 (1)

For each classification case, the ranks of the classifiers are shown in parentheses after the average F1 scores or the average ranks.

Table S4: *Mean F1 scores (F1) and standard deviations (SDs) of different three meta-models.*

Classifier	Mean F1 and SDs		
	Top Three Best Models (CatBoost + GBM + RF)	Three Diverse Performing Models (CatBoost + MLP + KNN)	Bottom Three Worst Models (LR + SVM + KNN)
Fatal Attacks	**0.831** ± 0.005	0.828 ± 0.004	0.815 ± 0.005
Injury Attacks	**0.730** ± 0.006	0.723 ± 0.006	0.701 ± 0.007
Property Damage Attacks	**0.823** ± 0.008	0.818 ± 0.007	0.801 ± 0.006

The bold numbers indicate the maximum values compared among different meta-models.

According to Table S4, the meta-models that combine the top three best classifiers (i.e., CatBoost, GBM, and RF) achieved the greatest mean F1 scores when compared to other meta-models. From the results, it can be suggested that the poorer performance of base models combined in a meta-model has, the lower value of the mean F1 score the meta-model achieves. This may be because meta-models are mainly based on the classification results obtained from base classifiers and discard the features of the original data. Thus, the more inaccurate predictions from the base classifiers combined in meta-models can negatively influence the classification performance of the meta-models.