Proceedings of International Structural Engineering and Construction, 8(1), 2021
Interdisciplinary Civil and Construction Engineering Projects
Edited by El Baradei, S., Abodonya, A., Singh, A., and Yazdani, S.

ISSN: 2644-108X
www.doi.org/10.14455/ISEC.2021.8(1).STR-53

MACHINE LEARNING FOR SEISMIC-INDUCED DAMAGE ESTIMATION OF STEEL TANKS

LAURA MICHELI[1], HAN Q. NGUYEN[2], LIN-CHING CHANG[2], and MAHMOUD FAYTAROUNI[3]

[1]*Dept of Civil and Environmental Engineering, Catholic University of America, Washington DC, USA*
[2]*Dept of Electrical Engineering and Computer Science, Catholic University of America, Washington DC, USA*
[3]*Genex Systems, Turner Fairbank Highway Research Center, McLean, USA*

Aboveground steel storage tanks are large vessels employed to store various liquids, including water, food, fertilizers, oil, and other hazardous chemicals. The damage and collapse of storage tanks can generate long-lasting consequences on built environment and communities. The seismic behavior of storage tanks is often evaluated employing fragility functions obtained simulating the tank under a variety of ground motion records. However, the computational demand of high-fidelity simulation models makes risk assessment a burdensome task. The use of data-driven surrogate models could represent a suitable solution to evaluate the seismic vulnerability of steel tanks rapidly. In this context, this paper presents an open dataset composed of 204 aboveground cylindrical steel liquid storage tanks with different geometric properties. The dataset was assembled based on past earthquake reconnaissance reports. In the dataset, different types of damage experienced by the steel tanks are divided into four classes, ranging from no damage to complete failure. Eight different machine learning algorithms are trained to predict the damage class of a steel tank as a function of its geometric properties and seismic excitation parameters. Stratified 5-fold cross-validation is used to split the dataset into training and testing subsets and to assess the prediction capability of the machine learning models. Results showed that the Support Vector Machine algorithm yielded the most accurate predictions, followed by Random Forest, XGBoost, and LightBoost. Overall, the paper demonstrated the feasibility of using machine learning models to predict the damage level of steel liquid storage tanks subjected to seismic hazard.

Keywords: Steel tanks, Damage classification, Machine learning, Seismic hazard.

1 INTRODUCTION

The damage exhibited by industrial facilities under recent earthquake events demonstrated that steel liquid storage tanks are vulnerable to seismic hazards. Damage and collapse of steel tanks could lead to undesired consequences, such as large-scale spillages of dangerous chemicals, fires, and unavailability of clean water in the wake of the earthquake. Modeling the behavior and predicting the damage of liquid storage tanks under seismic load is challenging due to the significant uncertainties inherent structural and hazard characteristics. High-fidelity numerical simulation models (e.g., finite element models) can be employed to realistically estimate the tank response under earthquake loads (Korkmaz *et al*. 2011, Lee *et al*. 2019). However, their high computational demand, combined with the large number of uncertainties to consider, makes risk assessment prohibitively time-consuming.

Surrogate models have been proposed to speed up risk assessment procedures (Micheli and Laflamme 2020). Surrogate models are simplified mathematical representations of a physical system and are usually employed as a replacement of the original, computationally expensive simulation model. For steel liquid storage tanks, many physics-based surrogate models are available in the literature. For example, Bakalis *et al*. (2017) employed a single-mass surrogate model characterized by elastic beam-elements and nonlinear springs to evaluate the seismic risk of steel liquid storage tanks. Later, Phan *et al*. (2019) utilized a simplified model consisting of an equivalent single-degree of freedom system to model the dynamic behavior of an unanchored steel liquid storage tank and perform seismic fragility analysis considering various uncertain model parameters. However, these analytical models are rarely informed by field data and lead to a relatively large computational demand when the number of uncertain variables to consider is large. This paper presents a preliminary investigation on the use of data-driven surrogate models for damage identification of steel tanks, where the machine learning (ML) algorithms are trained based on field earthquake reconnaissance data. Given the low computational demand of ML models, such an approach is suitable to rapidly assess the seismic risk of steel tanks, and it could be potentially integrated into a real-time decision-making framework for post-disaster management operations.

The objectives of this paper are: 1) to assemble a dataset of the damage reported by above-ground cylindrical steel liquid storage tanks under different earthquake events. Such dataset is derived from existing reconnaissance reports, and it includes 204 tanks. 2) To assess the capability of different ML algorithms to predict the damage class of steel liquid storage tanks as a function of tank proprieties and earthquake parameters. 3) To identify the classification algorithm most suitable to estimate the level of damage in steel liquid storage tanks. Such a classification model can potentially incorporate more reconnaissance data when available.

2 DATA-DRIVEN SURROGATE MODELS

2.1 Steel Liquid Tanks Dataset Description

A dataset of 204 above-ground, cylindrical, steel liquid storage tanks was assembled from past earthquake reconnaissance reports available in the literature (Cooper 1997). Both anchored and unanchored tanks were considered. For each tank, the data collected consists of tanks properties, earthquake characteristics, and corresponding seismic-induced damage. Tank's properties include diameter (D), height (H), thickness (t), and liquid height at the moment of the earthquake (H_L), while earthquake characteristics include magnitude (M) and peak ground acceleration (PGA), schematically represented in Figure 1 (a). For the majority of the tanks (158), the thickness was not reported in the reconnaissance report, and it was estimated based on the work of Cortes and Prinz (2017). Similarly, for some earthquake events, the PGA was estimated based on ground motion records available in the PEER database (Pacific Earthquake Engineering Research Center 2020) and the geographic location of the tank. Since the material properties of the steel tanks were not specified in the reconnaissance report, yielding stress and modulus of elasticity were assumed to be the same for all the tanks. The steel liquid storage tanks reported different types of damage, such as elephant foot buckling, roof damage, shell buckling, and complete collapse. Such damage types were grouped into four main classes (adapted from O'Rourke and So 2011): class 0, no damage; class 1, minor damage, including damage to the seal, minor roof damage, and in general, damage that requires simple repairs; class 2, severe damage, such as elephant foot buckling and shell buckling; and class 3, total failure, collapse of the tank, and widely spread damage that necessitates the replacement of the tank. The distribution of the steel tanks data among the four damage classes is shown in Figure 1 (b), along with a description of each class in Figure 1 (c). Note that Figure 1 (c) reports the types of damage experienced by the steel tanks available in the reconnaissance report (Cooper 1997). If more reconnaissance data becomes available, different damage modes may be considered, and the classes of damage types could be further refined. The dataset used in this study is publically available on GitHub (Micheli 2020).

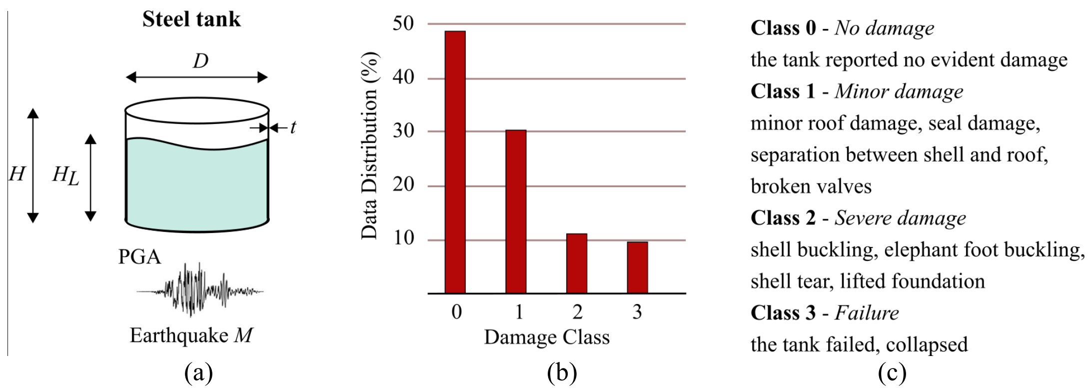

Figure 1. (a) Schematic representation of a steel liquid storage tank, where D is the tank diameter, H its height, t its thickness, H_L is the height of the liquid, M is the earthquake magnitude, and PGA is the peak ground acceleration; (b) distribution of the data among classes; (c) damage classes description.

Table 1. Statistics of the collected data.

	Input	Minimum	Maximum	Mean	Standard Dev.
Steel tank geometry	D	2 m	83 m	23 m	14 m
	H	3 m	45 m	11 m	4 m
	t	0.001 m	0.098 m	0.013 m	0.006 m
	H_L	0 m	45 m	10 m	9 m
Seismic hazard	M	5.0	8.5	7.0	0.6
	PGA	0.02 g	2.0 g	0.44 g	0.42 g

2.2 Machine Learning Algorithms

The steel liquid storage tanks dataset was employed to train and test different data-driven surrogate models (i.e., ML algorithms). Each surrogate model was constructed based on a set of N observations, $S = \{(\mathbf{x}^{(i)}, y^{(i)}), i = 1,\dots, N\}$, where $\mathbf{x}$ is the input vector and y is the output. In this study, the total number of observations available is $N = 204$. The inputs of the surrogate were selected as the tank's diameter (D), height (H), thickness (t), liquid height (H_L), earthquake magnitude (M), and peak ground acceleration (PGA). Table 1 reports the basic statistics of the inputs available in the dataset. Note that the material properties of the tanks were not taken as inputs because they were assumed to be identical for all the tanks, and therefore, would not add information to the data-driven models. The output y of the surrogate model was defined as the damage class (0, 1, 2, or 3). Stratified 5-fold cross-validation strategy (Viana *et al.* 2009) is used to split the dataset into training (size $n = 164$, 80% of N) and testing subsets (size $n_t = 40$, 20% of N) and to evaluate the ML models on the limited data samples. The training data were used to train the ML algorithms to estimate the damage class of a steel liquid storage tank as a function of the tanks properties (D, H, t, H_L) and earthquake parameters (M, PGA). The model performance on predicting the damage class in correspondence of the test data was used as an indication of the model performance on unseen data.

The ML algorithms investigated in this paper are briefly described in what follows. More information can be found in Mangalathu *et al.* (2020).

Decision Tree: it is a non-parametric ML model that estimates the value of a target variable by learning simple decision rules inferred from the input features. Two steps are usually necessary to create a decision tree: construction and pruning. Tree construction consists of dividing the training space into non-overlapping subspaces based on the information gain (i.e., Gini index or entropy). Tree pruning is then performed to prune the tree from some branches and avoid overfitting.

Random Forest: the model is composed of an ensemble of individual decision tree classifiers generated from the training set with bootstrap sampling. Each tree classifies a given input vector in a certain class, casting a vote for that class. The final prediction is selected as the class with most votes.

Support Vector Machine (SVM): it consists of constructing a set of hyperplanes that separate two or more classes. The optimal separation of classes is taken as the hyperplane that maximizes the margin distance (i.e., the sum of the distances to the hyperplane from the closest points of the classes) between data points belonging to two or more classes.

Naïve Bayes: it is a parametric classifier based on the Bayes theorem. It simplifies the learning process with the assumption that the features are independent from each other. In this study, a Gaussian distribution is assumed as prior distribution in the Bayes theorem.

AdaBoost, XGBoost, LightGBM Classifiers: the boosting classifiers enhance the performance of decision trees by generating new models that correct the errors of the previous trees. AdaBoost is an adaptive algorithm, where the training weights are updated based on the misclassified data from previous iterations. XGBoost and LightGBM are gradient boosting classifiers. The difference between these algorithms is in the rules they use to fit the strong classification model to the set of weak learners.

3 RESULTS AND DISCUSSION

The performance of the algoritims was assessed with several metrics, including accuracy, precision, and recall. As an example, the confusion matrices for three algorithms (i.e., SVM, Random Forest, XGBoost) from one of the folds are reported in Figure 2. Each element of the confusion matrix $C_{i,j}$ represents the number of observations known to be in class i and predicted to be in class j. Therefore, the diagonal elements (in green) are the classes (i.e., damage levels) correctly predicted by the model, while the off-diagonal elements (in red) are the categories classified incorrectly. The confusion matrices also report precision (percentage of damage classes correctly assigned by the model), recall (percentage of actual damage classes, or true positives, correctly assigned by the algorithm), and accuracy (number of classes guessed correctly divided by the total number of guesses). Precision and recall are local performance metrics related to each damage class, while accuracy is a global performance metric. A cross-comparison between the confusion matrices in Figure 2 shows that the SVM model led to the highest accuracy. Recall and precision highly depend on the considered damage class. SVM yielded the largest precision in classes 1 to 3, and the highest recall in classes 0, 2, and 3. Random Forest led to the largest precision in class 0 and recall in class 1, respectively.

Table 2 reports the performance of the ML algorithms averaged over the 5-fold cross-validation, including accuracy, recall (R), and precision (P) for each class (0, 1, 2, and 3). Results in Table 2 show that SVM yielded the highest accuracy (68.3%) on the testing data, followed by Random Forest (64.5%), XGBoost (62.0%), and LightBoost (62.0%). SVM also led to the highest values of recall and precision for most of the classes, except recall in class 0 and precision in classes 1 and 3. The results demonstrate that the non-parametric tree-based methods (i.e., Decision Tree, Random Forest) exhibited better performance than the parametric non-tree-based methods (i.e., Naïve Bayes). The gradient boosting algorithms (i.e., XGBoost, AdaBoost, LightBoost) led to a similar performance in terms of accuracy. Overall, the results highlight the importance of considering different metrics when evaluating the performance of different ML algorithms at predicting damage classes in steel tanks.

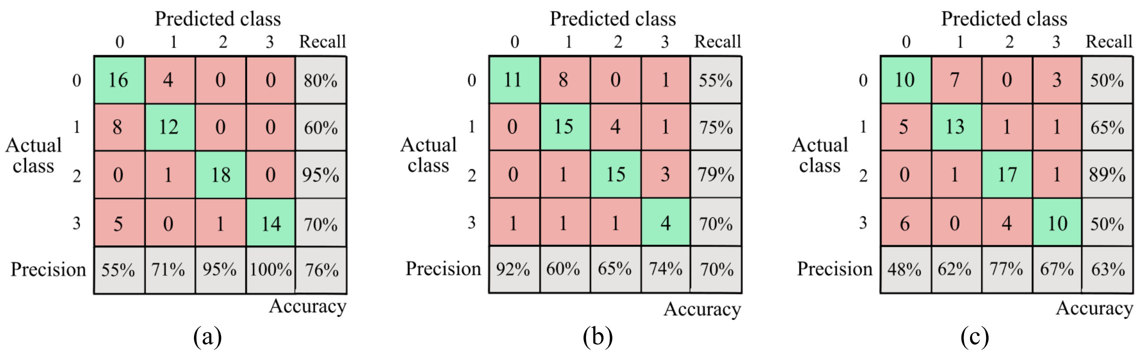

Figure 2. Confusion matrix using one fold of the test set: (a) SVM; (b) Random Forest; (c) XGBoost.

Table 2. Performance comparison of the machine learning algorithms (averaged over the 5 folds).

Algorithm	Accuracy (%)	R0 (%)	P0 (%)	R1 (%)	P1 (%)	R2 (%)	P2 (%)	R3 (%)	P3 (%)
Decision Tree	58.2	47.0	27.3	55.4	42.4	65.9	85.9	51.3	67.7
Random Forest	64.5	84.6	37.5	60.9	75.8	66.9	83.8	67.5	79.9
SVM	68.3	55.9	65.7	69.9	64.6	84.1	88.9	92.5	72.6
Naïve Bayes	54.4	44.4	29.1	54.3	38.0	65.7	87.4	52.5	67.5
XGBoost	62.0	50.1	39.7	58.5	74.7	72.1	82.9	72.5	61.9
AdaBoost	58.2	84.6	37.5	60.9	75.8	66.9	83.8	67.5	70.9
LightGBM	62.0	73.7	31.5	56.7	68.6	67.9	86.8	67.7	66.0

4 CONCLUSIONS

This paper presented the preliminary results of an ongoing investigation on the use of ML models for predicting the damage experienced by steel liquid storage tanks under seismic hazard. A dataset of 204 above-ground cylindrical steel liquid storage tanks characterized by different geometric properties and types of seismic-induced damage was first assembled. Then, eight ML algorithms were trained based on part of the data (training dataset), and their performance at predicting the damage class of steel liquid storage tanks was assessed on the remaining data (testing dataset) using stratified 5-fold cross-validation. Results highlighted the necessity to evaluate different algorithms before selecting a ML model for damage classification of steel liquid storage tanks. Based on its performance, the SVM appeared to be the most suitable ML algorithm to predict the damage class of steel liquid storage tanks.

Overall, this work demonstrated the capacity of ML models to predict the damage class of steel liquid storage tanks. Note that the proposed algorithm can be used only for tanks with geometric properties and earthquake parameters lying in the range reported in Table 1. Furthermore, the distribution of the data between damage classes (Figure 1 (b)) is imbalanced and more data from classes 2 and 3 are needed to improve the accuracy of the ML model. Finally, the

ML model can predict only the damage class of a certain steel liquid storage tank and not the exact damage type (e.g., elephant foot buckling). Future work entails improving the performance of the ML algorithms by providing a more balanced dataset, applying other techniques to handle the issue of imbalance data, exploring a larger set of inputs, and integrating the model into a vulnerability assessment framework where the ML algorithm rapidly estimates the damage class of a steel tank as a function of its characteristics and earthquake parameters.

References

Bakalis, K., Vamvatsikos, D., and Fragiadakis, M., *Seismic Risk Assessment of Liquid Storage Tanks via Nonlinear Surrogate Model,* Earthquake Engineering and Structural Dynamics, 46(15), 2851-2868, July, 2017.

Cooper, T. W., *A Study of the Performance of Petroleum Storage Tanks During Earthquakes, 1933-1995,* US National Institute of Standards and Technology, 1997.

Cortes, G., and Prinz, G. S., *Seismic Fragility Analysis of Large Unanchored Steel Tanks Considering Local Instability and Fatigue Damage*, Bulletin of Earthquake Engineering, 15(3), 1279-1295, August, 2017.

Korkmaz, K. A., Sari, A., and Carhoglu, A. I., *Seismic Risk Assessment of Storage Tanks in Turkish Industrial Facilities,* Journal of Loss Prevention in the Process Industries, 24(4), 314-320, July, 2011.

Lee, S., Kim, B., and Lee, Y. J., *Seismic Fragility Analysis of Steel Liquid Storage Tanks Using Earthquake Ground Motions Recorded in Korea*, Mathematical Problems in Engineering, July, 2019.

Mangalathu, S., Jang, H., Hwang, S. H., and Jeon, J. S., *Data-Driven Machine-Learning-Based Seismic Failure Mode Identification of Reinforced Concrete Shear Walls,* Engineering Structures, 208, 110331, April, 2020.

Micheli, L., *GitHub*, 2020, Retrieved from https://github.com/lauramch/Steel-Liquid-Storage-Tanks-Database on June 04, 2021.

Micheli, L., and Laflamme, S., *Kriging-Based Design for Robust High-Performance Control Systems,* ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 6(4), 04020037, December, 2020.

O'Rourke, M. J., and So, P., *Seismic Fragility Curves for On-Grade Steel Tanks,* Earthquake Spectra, 16(4), 801-815, November, 2000.

Pacific Earthquake Engineering Research Center, *PEER Ground Motion Database*, 2020, retrieved from https://ngawest2.berkeley.edu/.

Phan, H. N., Paolacci, F., and Alessandri, S., *Enhanced Seismic Fragility Analysis of Unanchored Steel Storage Tanks Accounting for Uncertain Modeling Parameters,* Journal of Pressure Vessel Technology, 141(1), December, 2019.

Viana, F. A., Haftka, R. T., and Steffen, V., *Multiple Surrogates: How Cross-Validation Errors Can Help Us to Obtain the Best Predictor,* Structural and Multidisciplinary Optimization, 39(4), 439-457, January, 2009.