Available online at www.eccomasproceedia.org
Eccomas Proceedia COMPDYN (2025) 219-228

ECCOMAS
Proceedia

COMPDYN 2025
10th ECCOMAS Thematic Conference on
Computational Methods in Structural Dynamics and Earthquake Engineering
M. Papadrakakis, M. Fragiadakis (eds.)
Rhodes Island, Greece, 15-18 June 2025

PHYSICS-INFORMED MACHINE LEARNING MODEL FOR PREDICTING THE DISPLACEMENT DEMANDS OF STRUCTURES: INSIGHTS FROM THE 2023 KAHRAMANMARAŞ EARTHQUAKES

Furkan Çalım[1*], Ahmet Güllü[1], and Ercan Yüksel[1]

[1] Faculty of Civil Engineering, Istanbul Technical University
Maslak, 34469 Sariyer, Istanbul, Türkiye
calimf@itu.edu.tr, ahmetgullu@itu.edu.tr, yukselerc@itu.edu.tr

Abstract

Accurate determination of displacement demands from ground excitations is essential for evaluating the seismic performance of existing buildings and designing new structures resilient to natural and man-made disasters. Energy-based evaluation and design approaches have emerged as effective tools for achieving this objective. This study investigates the correlation between seismic input energy and top displacement demands, laying the groundwork for a reliable methodology to predict displacement demands in structural systems, focusing on the 2023 Kahramanmaraş earthquake sequence. Response history analyses were performed on single-degree-of-freedom systems using various ground motion records, and the relationships were examined through parametric studies involving vibrational period and damping ratio. Building on these findings, a novel machine-learning model employing the XGBoost algorithm was developed to predict the relationship between seismic input energy and top displacement demands. The XGBoost-based approach demonstrated enhanced predictive accuracy, providing a robust tool for estimating structural demands, particularly top displacement demands, and contributing to seismic risk assessment and mitigation efforts.

Keywords: Seismic Input Energy, Top Displacement Demand, Machine Learning, XGBoost, Predictive Modeling, Kahramanmaraş Earthquake Sequence.

doi: 10.7712/120125.12403.25934

1 INTRODUCTION

Accurate prediction of displacement demands under seismic loading is crucial for the seismic performance evaluation and design of nonlinear structural systems. Traditional nonlinear time history analyses are accurate yet computationally demanding, leading to the development of simplified nonlinear static procedures (NSP) that require target displacement estimation [1-3]. The equal displacement principle [4] forms the basis of many NSP methods, though variations in inelastic displacement ratios due to ground motion characteristics and structural parameters have been studied extensively.

Energy-based procedures have gained popularity as an alternative to conventional methods for predicting the seismic response of nonlinear structural systems (i.e., especially displacement demands) since they consider key parameters like the frequency content, duration, and cumulative damage potential of ground motions [5-6].

On the other hand, recent advancements in machine learning techniques have opened new avenues for enhancing predictive models in earthquake engineering. Among various machine learning techniques, ensemble learning methods (e.g., support vector machines [7], convolutional neural networks [8], random forest [9], and XGBoost [10]) have gained significant attention due to their ability to model complex nonlinear relationships with high accuracy.

This study aims to bridge the gap between energy-based seismic evaluation methods and modern machine learning approaches by developing a physics-informed model with an XGBoost algorithm to improve the prediction of displacement demands from ground excitations. A large dataset of ground motions was created by performing extensive linear and nonlinear time history analyses on SDOF systems. The computed seismic input energy demands, along with several other intensity measures, were employed as input features for the XGBoost model. In the end, the predictive performance of the model was investigated regarding the root mean squared error, coefficient of determination, feature importance, and learning curve.

2 DATASET AND FEATURE ENGINEERING

A comprehensive and well-curated dataset is needed to develop a robust physics-informed machine learning model for predicting structures' displacement demands. This section describes the process of collecting and preparing ground motion data, generating structural response data through time history analyses, and selecting relevant features for the model.

2.1 Ground motion data

On February 6th, 2023, an earthquake with a moment magnitude of 7.8 occurred in Pazarcık county of Kahramanmaraş. Only after nine hours, the subsequent ground excitation with a moment magnitude of 7.6 was recorded in Elbistan County. The earthquakes generated strong ground shaking, resulting in catastrophic human and economic loss across south-central Türkiye and northwest Syria. The earthquake sequence was recorded as one of the strongest in the history of this seismically active area.

The Pazarcık Earthquake, the first event in the sequence, was selected for this study due to its strong ground motion records. For the dataset generation, a total of 200 accelerometer stations were selected from the earthquake-affected region. Utilizing both horizontal components of the accelerometer stations yielded 400 separate ground motion records for the analyses.

2.2 Response data

The displacement demand for each ground acceleration record was obtained through a set of linear and nonlinear time history analyses on single-degree-of-freedom (SDOF) systems

[11]. The numerical analyses were conducted using Python-based simulations for the systems with varying natural periods and ductility levels (i.e., 1, 2, 4, 6, and 8). The objective of the nonlinear analyses was to achieve a target ductility, ensuring inelastic deformation without significant strength loss. Numerical methods were used to solve nonlinear differential equations, employing the piecewise exact method by Aydınoğlu and Fahjan [12]. The post-yield stiffness ratio and stability coefficient were set to zero, resulting in an elastic-perfectly-plastic hysteresis model.

2.3 Feature selection

Feature selection is a critical step in developing a robust machine learning model, as it directly influences the model's performance. To develop a physics-informed model for displacement demand predictions, a set of intensity measures were selected as input features, as summarized in Table 1. In addition to the intensity measures, the seismic input energy, ductility ratio, impulsiveness, and significant duration were included as input features. Since the displacement demand exhibited a linear relationship with the square root of seismic input energy, this physics-informed feature was incorporated into the model.

Table 1: Intensity measures selected as input features.

Notation	Definition	Formulation	Notation	Definition	Formulation
PGA	Peak ground acceleration	$PGA = \max\left(\left\|\ddot{u}_g(t)\right\|\right)$	d_{RMS}	Root mean square of ground displacement [17]	$d_{RMS} = \sqrt{\frac{1}{t}\int_0^t \left[u(t)\right]^2 dt}$
PGV	Peak ground velocity	$PGV = \max\left(\left\|\dot{u}_g(t)\right\|\right)$	I_c	Characteristic intensity [18]	$I_C = a_{RMS}^{1.5}\sqrt{t}$
PGD	Peak ground displacement	$PGD = \max\left(\left\|u_g(t)\right\|\right)$	*CAV*	Cumulative absolute velocity [19]	$CAV = \int_0^{t_f}\left\|\ddot{u}_g(t)dt\right\|$
I_F	Fajfar intensity [13]	$I_F = PGV \cdot t_d^{0.25}$	*CAD*	Cumulative absolute displacement [20]	$CAD = \int_0^{t_f}\left\|\dot{u}_g(t)dt\right\|$
I_a	Compound acceleration-related intensity [14]	$I_a = PGA \cdot t_d^{1/3}$	*SED*	Specific energy density [21]	$SED = \int_0^{t_f}\dot{u}_g^2(t)dt$
I_v	Compound velocity-related intensity [14]	$I_v = PGV^{2/3} \cdot t_d^{1/3}$	I_H	Housner intensity [22]	$I_H = \int_{0.1}^{2.5} S_{pv}dT$
I_d	Compound displacement-related intensity [14]	$I_d = PGD \cdot t_d^{1/3}$	*ASI*	Acceleration spectrum intensity [23]	$ASI = \int_{0.1}^{0.5} S_a dT$
VA	*PGV*-to-*PGA* ratio [15]	$VA = PGV/PGA$	*VSI*	Velocity spectrum intensity [23]	$VSI = \int_{0.1}^{2.5} S_v dT$
I_A	Arias intensity [16]	$I_a = \frac{\pi}{2g}\int_0^{t_d}\ddot{u}^2(t)dt$	$V_{EIa}SI$	Absolute input equivalent velocity spectrum intensity [24]	$V_{EIa}SI = \int_{0.1}^{3.0}\sqrt{2E_{Ia}}dT$
a_{RMS}	Root mean square of ground acceleration [17]	$a_{RMS} = \sqrt{\frac{1}{t}\int_0^t\left[\ddot{u}(t)\right]^2 dt}$	$V_{EIr}SI$	Relative input equivalent velocity spectrum intensity [24]	$V_{EIr}SI = \int_{0.1}^{3.0}\sqrt{2E_{Ir}}dT$
v_{RMS}	Root mean square of ground velocity [17]	$v_{RMS} = \sqrt{\frac{1}{t}\int_0^t\left[\dot{u}(t)\right]^2 dt}$	*ESI*	Energy spectral intensity [25]	$ESI = \int_{0.1}^{2.5} S_{IEP}(T,\xi)dT$

Note: S_v: 5% damped acceleration spectrum; S_v: 5% damped velocity spectrum; S_{pv}: 5% damped pseudo-velocity spectrum; $t_d = t_2 - t_1$ where t_2 = t(95%I_A) and t_1 = t(5%I_A); ξ: Damping ratio

3 METHODOLOGY

A physics-informed machine learning model was developed to predict the displacement demands of structures. The employed methodology consists of three main components: defining the machine learning model, dataset training and testing strategy, and model evaluation.

3.1 Machine learning model

The XGBoost (Extreme Gradient Boosting) algorithm is a powerful and scalable learning technique based on gradient boosting frameworks. It is a collective learning technique that combines the predictions of multiple weak learners to create a strong predictive model. The key idea behind that is to iteratively improve the model by focusing on the errors made by previous learners. It is widely used for regression, classification, and ranking tasks due to its high performance, flexibility, and ability to handle large datasets efficiently.

An XGBoost regressor was initialized with the following hyperparameters, which significantly influence the model performance. Here, the "*objective*" is the loss function to be optimized in regression tasks. The selected objective loss function specifies that the model should minimize the squared error between the predicted values and the actual target values. The "*n_estimators*" is the number of boosting rounds (trees) to be trained. A higher number improves model performance but increases computational cost. The "*learning_rate*" is the step size at each iteration. A lower learning rate requires more trees but can lead to better generalization. The "*max_depth*" is the maximum depth of each tree. Deeper trees can capture more complex patterns but may overfit the data. Finally, the "*random_state*" is a seed value to ensure reproducibility of results.

The process starts with an initial model that predicts the target variable. The residuals (i.e., errors) between the actual and predicted values are computed. At each iteration, a new weak learner (i.e., decision tree) is trained to predict the residuals of the previous model. The predictions of the new tree are added to the existing model, gradually reducing the residuals. The algorithm minimizes the loss function (i.e., mean squared error in this specific study) using gradient descent. After a specified number of iterations, which is controlled by the "*n_estimators*" parameter, the final model is obtained by summing the predictions of all weak learners.

```
model = xgb.XGBRegressor(
    objective="reg:squarederror",     # Loss function for regression
    n_estimators=100,                 # Number of boosting rounds
    learning_rate=0.1,                # Step size for each boosting iteration
    max_depth=6,                      # Maximum depth of a tree
    random_state=42                   # Seed for reproducibility
)
```

3.2 Training and testing strategy

The dataset, which includes the features (i.e., seismic input energy demand, impulsivity, ductility ratio, and several intensity measures) and the target variable (i.e., top displacement demand), was loaded into the developed Python-based code. Later, it was randomly split into 80% training and 20% testing sets to ensure the robust evaluation of the model. Data normalization was also applied to improve numerical stability and model performance. After training the model on the normalized training data, it was used to predict displacement demands on the test set.

```
data = pd.read_excel("seismic_data.xlsx", sheet_name="Sheet1")
features = data.columns.tolist()
features.remove("Displacement Demand")
target = "Displacement Demand"
x = data[features]
y = data[target]

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)

scaler = StandardScaler()
x_train = scaler.fit_transform(x_train)
x_test = scaler.transform(x_test)

model.fit(x_train, y_train)
y_pred = model.predict(x_test)
```

3.3 Model evaluation

The performance of the XGBoost-based machine learning model was evaluated using two metrics, i.e., the root mean squared error (RMSE) and the coefficient of determination (R^2). The RMSE measures the average deviation of predicted values from actual values, while the R^2 score quantifies the proportion of variance in the target variable explained by the model. In addition, the importance of each feature was analyzed to gain insights into the developed model's decision-making process. It identifies the most influential features in predicting the target value. Finally, the learning curve was calculated to assess the model's generalization ability and potential for improvement with more training data.

```
rmse = mean_squared_error(y_test, y_pred, squared=False)
r2 = r2_score(y_test, y_pred)

importance = model.feature_importances_
feature_importance = pd.DataFrame({'Feature': features, 'Importance':
importance}).sort_values(by='Importance', ascending=False)

train_sizes, train_scores, test_scores = learning_curve(model, x, y, cv=5)
train_scores_mean = train_scores.mean(axis=1)
test_scores_mean = test_scores.mean(axis=1)
```

4 RESULTS AND DISCUSSION

Figure 1 presents the feature correlation matrix, which provides insight into the relationships between several intensity measures and the target variable (i.e., top displacement demand). In this matrix, each cell represents the Pearson correlation coefficient [26] between two variables ranging from -1 (strong negative correlation) to +1 (strong positive correlation). The seismic input energy exhibited the highest correlation coefficient with the top displacement demand, which suggests that it has the most significant linear correlation with the target value in the model.

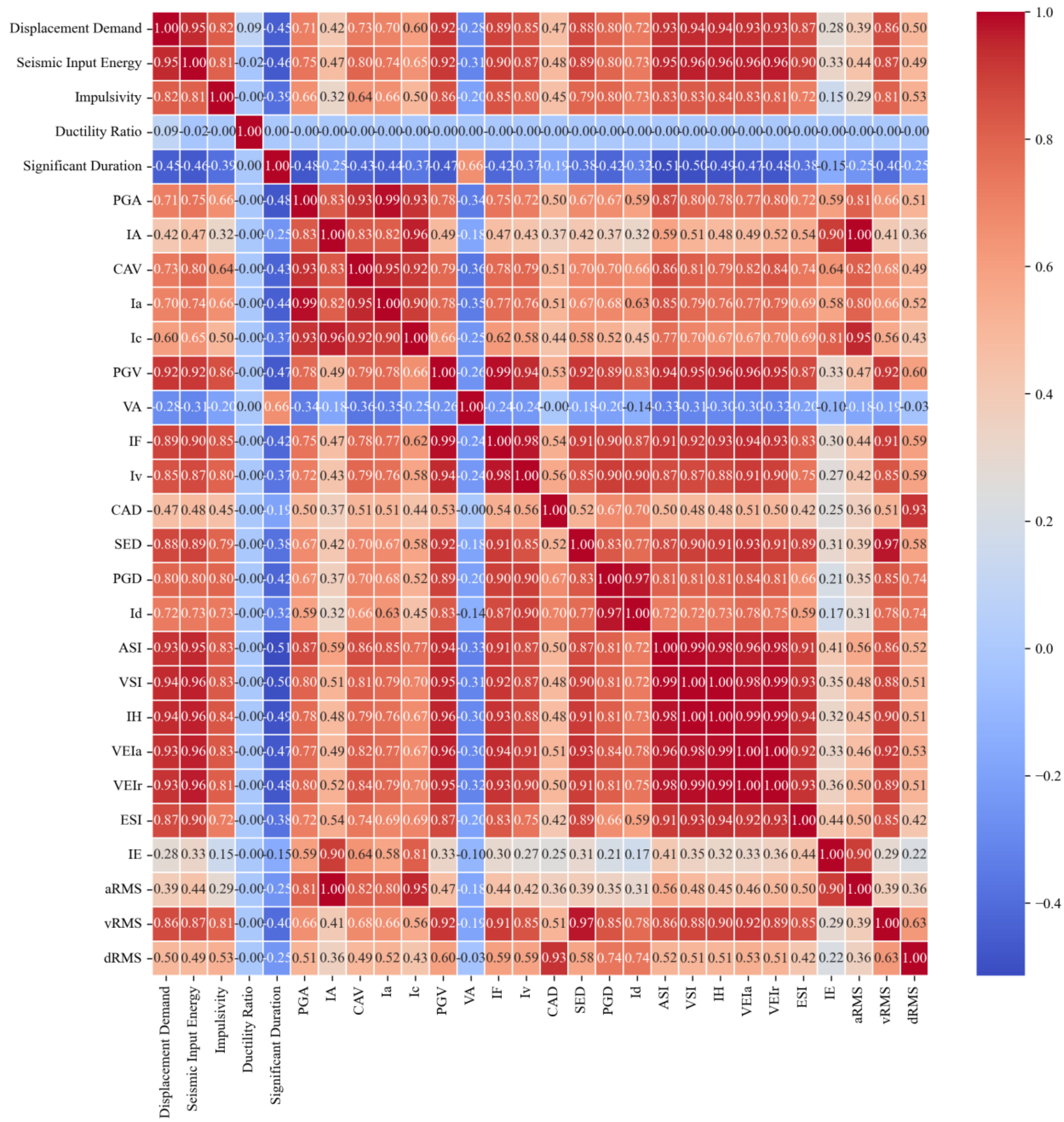

Figure 1: Feature correlation matrix (including the target).

The future importance values in the developed XGBoost model are displayed in Figure 2. They are automatically computed during model training based on one of several methods, e.g., weight, gain, and cover. By default, XGBoost uses the gain method (i.e., the average improvement in accuracy brought by the feature when used in splits) to measure the contribution of each feature to the model's performance. Features with higher importance values have a stronger influence on the model's predictions, while those with lower values contribute less. For the specific case in this study, the seismic input energy demand resulted in the highest importance value (i.e., 0.798) among all other features. While *VSI* (velocity spectrum intensity), *ASI* (acceleration spectrum intensity), and I_H (Housner intensity) yielded feature importance values of 0.125, 0.024, and 0.019, respectively, all other intensity measures contributed minimally, with importance values below 0.010.

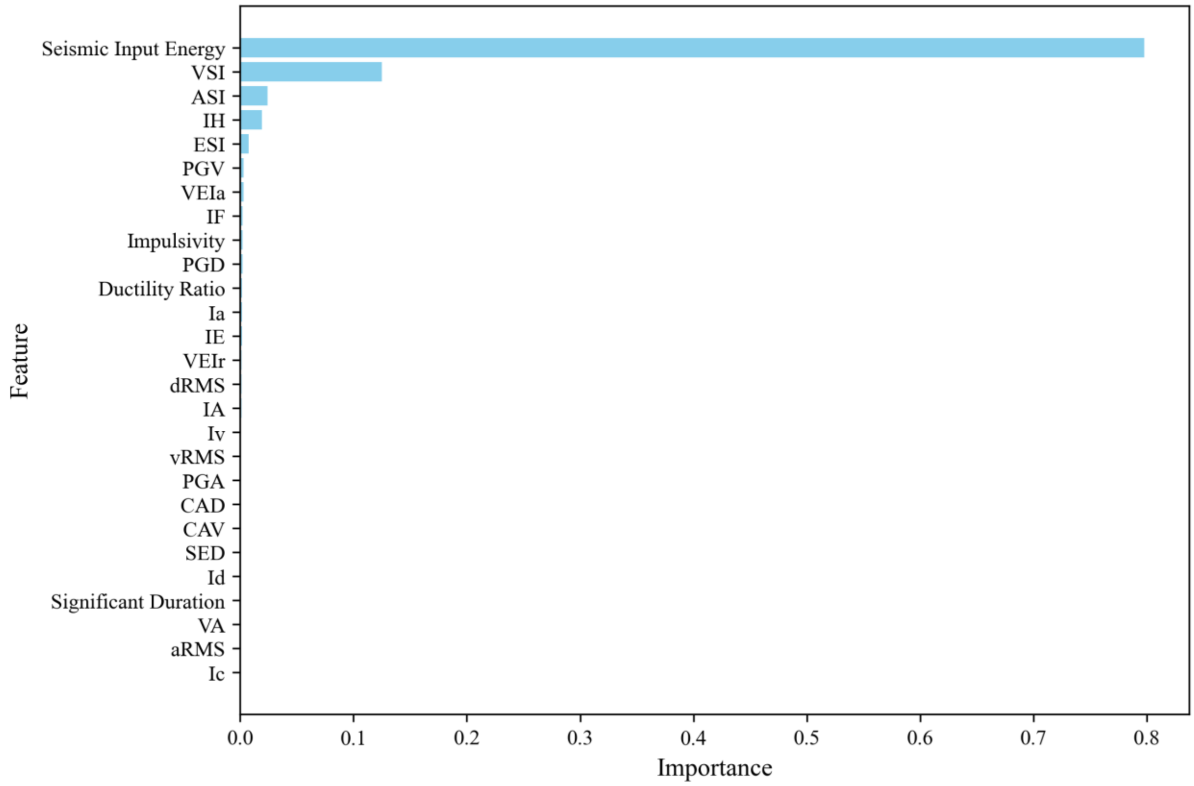

Figure 2: Feature importance in the XGBoost model.

The predicted versus actual values for the displacement demands are presented in Figure 3 for the testing set (i.e., 400 data points). Each point represents data obtained from an individual ground motion record. The R^2 value, which is a statistical measure of the goodness of fit of the regression model, for the predicted versus actual values for the displacement demands was computed as 0.9647. On the other hand, the root mean squared error was calculated as 0.0150. These results suggest that the model fits the data very well and has strong predictive accuracy.

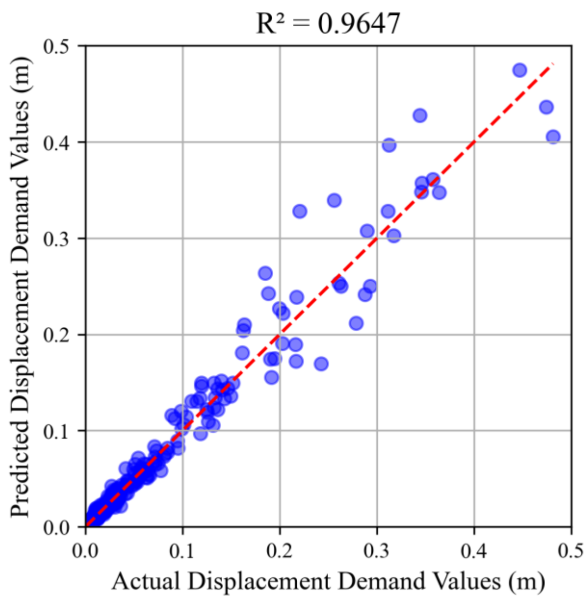

Figure 3: Predicted versus actual values for the displacement demands.

Figure 4 shows the generated machine learning model's learning curve, a graphical representation showing how performance improves as the training dataset size increases. The R^2 score is displayed in the graph for both the training and validation sets. The dataset consists of a total of 2000 data points, with 1600 used for training and 400 reserved for testing. The learning curve reveals that the validation score gradually increases with larger training set sizes, indicating that the model benefits from more data. Also, the gap between the training and validation scores suggests that the model has not yet fully converged. Thus, further increasing the training set size could result in the validation score approaching the training score, and it could improve the model's prediction performance.

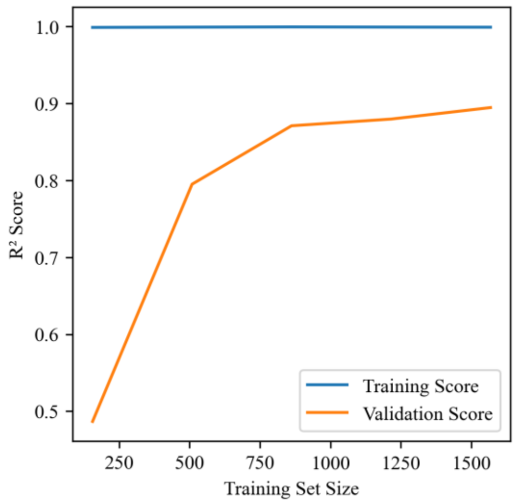

Figure 4: Learning curve.

5 CONCLUSIONS

In this study, a physics-informed machine learning model was successfully developed using the XGBoost algorithm to predict displacement demands, with a special focus on the 2023 Kahramanmaraş Earthquakes. As a result, the following conclusions could be driven:

- The XGBoost model demonstrated exceptional performance by achieving an R^2 value of 0.9647 on the testing dataset. This high predictive accuracy proves the generated machine learning model's robustness and reliability for seismic performance evaluation and design applications.
- The feature importance analysis concluded that the seismic input energy was the most influential feature in comparison with all other investigated intensity measures. This finding aligns with the feature correlation matrix, which showed that seismic input energy had the strongest linear correlation (i.e., correlation coefficient of 0.95) with the displacement demand.
- The learning curve demonstrated the generated model's potential for improvement with additional training data, as the validation score gradually approached the training score with increasing training set size.
- Future work could explore expanding the dataset to analyze numerous historical earthquake events, optimizing the hyperparameters for better model performance, and evaluating the success of some other machine learning approaches (e.g., random forest, support vector machines, and convolutional neural networks).

This study contributes significantly to the literature on energy-based performance evaluation and seismic design. It bridges the gap between energy-based methods and modern machine learning techniques, demonstrating their combined potential. The developed physics-informed machine learning model provides an accurate tool for predicting the critical displacement demands of structural systems, establishing a foundation for future research in this field.

ACKNOWLEDGMENTS

This study has been supported by TÜBİTAK (The Scientific and Technological Research Council of Türkiye) within the scope of the 2224-A International Scientific Events Participation Support Program.

REFERENCES

[1] ATC-40, Seismic evaluation and retrofit of concrete buildings, Applied Technology Council, Volume 1, Redwood City, CA, 1996.

[2] FEMA 356, Prestandard and Commentary for the Seismic Rehabilitation of Buildings. Washington, D.C.:Federal Emergency Management Agency, 2000.

[3] P. Fajfar, A nonlinear analysis method for performance-based seismic design. *Earthquake Spectra* **16(3)**, 573–592, 2000. https://doi.org/10.1193/1.1586128

[4] A. Veletsos, N. Newmark, Effect of inelastic behavior on the response of simple systems to earthquake motion. *2nd World Conference on Earthquake Engineering*, Tokyo, Japan, 1960.

[5] K. Fujii, M. Shioda, Energy-based prediction of the peak and cumulative response of a reinforced concrete building with steel damper columns. *Buildings* **13(2)**, 401. https://doi.org/10.3390/buildings13020401

[6] A. Güllü, F. Çalım, E. Yüksel, Estimation of the story response parameters through the seismic input energy for moment-resisting frames. *Soil Dynamics and Earthquake Engineering* **164**, 107636, 2023. https://doi.org/10.1016/j.soildyn.2022.107636

[7] C. Cortes, V. Vapnik, Support-vector networks. *Machine Learning*, **20(3)**, 273–297, 1995. https://doi.org/10.1007/BF00994018

[8] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86(11)**, 2278–2324, 1998. https://doi.org/10.1109/5.726791

[9] L. Breiman, Random forests. *Machine Learning*, **45(1)**, 5–32, 2001. https://doi.org/10.1023/A:1010933404324

[10] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*, San Francisco, CA, USA, pp. 785–794, August 13-17, 2016. https://doi.org/10.1145/2939672.2939785

[11] F. Çalım, E. Yüksel, Energy and displacement demand correlations for the 2023 Kahramanmaras Türkiye Earthquake Sequence. *3rd International Workshop on Energy-Based Seismic Engineering (EBSE 2025)*, Istanbul, Türkiye, 2025.

[12] M.N. Aydınoğlu, Y.M. Fahjan, A unified formulation of the piecewise exact method for inelastic seismic demand analysis including the P‐delta effect. *Earthquake Engineering and Structural Dynamics* **32(6)**, 871–890, 2003. https://doi.org/10.1002/eqe.252

[13] P. Fajfar, T. Vidic, M. Fischinger, A measure of earthquake motion capacity to damage medium-period structures. *Soil Dynamics and Earthquake Engineering* **9(5),** 236–242, 1990. https://doi.org/10.1016/s0267-7261(05)80002-8

[14] R. Riddell, J.E. Garcia, Hysteretic energy spectrum and damage control. *Earthquake Engineering and Structural Dynamics* **30(12)**, 1791–1816, 2001. https://doi.org/10.1002/eqe.93

[15] H. Sucuoğlu, A. Nurtuğ, Earthquake ground motion characteristics and seismic energy dissipation. *Earthquake Engineering and Structural Dynamics* **24(9)**, 1195–1213, 1995. https://doi.org/10.1002/eqe.4290240903

[16] A. Arias, A measure of earthquake intensity. In: R.J. Hansen (ed.), Seismic Design for Nuclear Power Plants, MIT Press, 438-483, 1970, Cambridge.

[17] M. McCann, RMS acceleration and duration of strong ground motion. The Jhon A Blume Earthquake Engineering Center, Report No: 46, 1980, Stanford University, CA.

[18] Y.J. Park, A.H.S. Ang, Y.K. Wen, Seismic damage analysis of reinforced concrete buildings. *Journal of Structural Engineering* **111(4)**, 740–757, 1985. https://doi.org/10.1061/(asce)0733-9445(1985)111:4(740)

[19] Electrical Power Research Institute (EPRI). A criterion for determining exceedence of the operating basis earthquake. EPRI NP-5930, Palo Alto, 1988.

[20] K. Mackie, B. Stojadinovic, Seismic demands for performance-based design of bridges. PEER Report, 2003.

[21] S.K. Sarma, Energy flux of strong earthquakes. *Tectonophysics* **11(3)**, 159–173, 1971. https://doi.org/10.1016/0040-1951(71)90028-x

[22] G.W. Housner, Spectrum intensities of strong motion earthquakes. *Proceedings Symposium on Earthquake and Blast Effect on Structures*. Earthquake Engineering Research Institute, Los Angeles (CA), 1952.

[23] J. Von Thun, L. Roehm, G. Scott, J. Wilson. Earthquake ground motions for design and analysis of dams. *Earthquake Engineering and Soil Dynamics II—Recent Advances in Ground-Motion Evaluation, Geotechnical Special Publication* **20**, 463–481, 1988.

[24] F. Mollaioli, A. Lucchini, Y. Cheng, G. Monti. Intensity measures for the seismic response prediction of base-isolated buildings. *Bulletin of Earthquake Engineering* **11(5)**, 1841–1866, 2013. https://doi.org/10.1007/s10518-013-9431-x

[25] C. Frau, S. Panella, M. Tornello, Input energy spectra for pulse-like ground motions. In: H. Varum, A. Benavent-Climent, F. Mollaioli (eds.), Energy-Based Seismic Engineering. IWEBSE 2023. Lecture Notes in Civil Engineering, Springer Cham, 216–233, 2023. https://doi.org/10.1007/978-3-031-36562-1_17

[26] K. Pearson, Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 1895. https://doi.org/10.1098/rspl.1895.0041