ACI STRUCTURAL JOURNAL TECHNICAL PAPER

Title No. 121-S06

A Framework to Set Performance Requirements for Structural Component Models: Application to Reinforced Concrete Wall Shear Strength

by Matías Rojas-León, Saman A. Abdullah, Kristijan Kolozvari, and John W. Wallace

Numerous models to predict the shear strength of reinforced concrete structural walls have been proposed in the literature. Evaluation of the predictive performance of new models relative to existing models is often challenging because the models were created with different levels of complexity and calibrated using different databases. More complex models are expected to have less variance than simpler models, and target performance metrics for models of different complexity do not exist. In addition, a common, comprehensive database should be used to enable direct comparisons between different models. To address these issues, the present study applies statistical and machine-learning approaches to propose a five-step framework to establish target performance metrics for models with different levels of complexity. Application of the framework is demonstrated by addressing the problem of estimating wall shear strength using a comprehensive database of 340 shear-controlled wall tests.

Keywords: machine learning; model performance; statistics; structural wall; wall shear.

INTRODUCTION

Over the last several years, researchers have assembled comprehensive component databases to enable the development of more complex capacity (for example, stiffness, strength, and deformation) models using more sophisticated statistical and machine-learning (ML) approaches. Evaluating and comparing the performance of different capacity models proposed in the literature is often challenging because: a) they were developed using different databases and a model may have substantially different performance (bias, variance) when evaluated against a different database; b) more complex models are expected to have less variance than a less-complex model—however, target performance metrics for models of different complexity do not exist; and c) optimal model performance is often not studied, so it is unknown whether a model with better performance is possible. In addition, many existing models were calibrated using relatively small databases—for example, less than 100 or so tests—such that insufficient data existed to properly train and test model performance, or training and testing were not even considered as part of the model development process.

To address these challenges, a framework is proposed to apply statistical and ML approaches to establish target performance requirements for models with different levels of complexity based on the use of a common, comprehensive database. Application of the proposed framework requires training of ML models to establish specific model performance requirements, where target errors are expressed in terms of the mean value and coefficient of variation (COV) of the true-to-predicted ratios. Once these metrics have been established, an additional study is required to develop a model that meets these requirements; this additional step is not addressed in this paper.

The methodology is demonstrated by addressing the problem of assessing wall shear strength using a comprehensive database of 340 walls reported to have failed in shear. This database was extracted from a larger database of more than 1100 tests collected from more than 250 experimental programs recently compiled by Abdullah and Wallace (2018, 2021) and Abdullah (2019). This application was picked because the wall shear strength equation in ACI 318-19 has remained essentially unchanged for the last 60 years despite a significant number of models being published in the literature. Although most of the published models (equations) are similar in complexity, significant model variance was noted when the models were assessed against a database that was different from the one used to develop and calibrate a given model (Gulec et al. 2009; Sánchez-Alejandre and Alcocer 2010; Carrillo and Alcocer 2013; Kassem 2015). These issues arise because the databases typically are of different sizes (number of tests), do not include the same wall tests, and have different ranges of parameters (for example, walls with rectangular cross sections versus walls with rectangular and flanged cross sections). Most of the studies also did not address the trade-off between underfitting versus overfitting (Höge et al. 2018) to examine the possibility that a model of equivalent complexity might have better predictive performance. The number of tests included in the wall shear database (340) and the number of variables for each test are expected to be typical of engineering problems that would benefit from the proposed methodology. Finally, none of the models met the set of performance requirements established in this paper for the given level of model complexity.

ACI Structural Journal, V. 121, No. 1, January 2024.

MS No. S-2022-407.R1, doi: 10.14359/51739186, received August 1, 2023, and reviewed under Institute publication policies.

RESEARCH SIGNIFICANCE

Evaluating and comparing the performance of models used to estimate structural component capacities is often challenging because the models were created with different levels of complexity and calibrated using databases with different numbers of tests and parameters. In addition, sufficient data may not have existed to properly train and test model performance, or training and testing were not considered. To address these challenges, a framework is proposed to apply statistical and ML approaches to establish model performance requirements for models of different complexities by training ML models to establish target errors expressed in terms of the mean value and COV of the true-to-predicted ratios. Application of the proposed framework is demonstrated by assessing the problem of estimating reinforced concrete (RC) wall shear strength.

REVIEW OF EXISTING WALL SHEAR STRENGTH MODELS

Models calibrated using statistical inference

Rojas-León (2022) presented a detailed literature review of existing models used in building codes and standards to predict the shear strength of RC walls (Appendix A,* Table A.1). The review reveals that all models use a $V_n = V_c + V_s$ format, where V_c and V_s are the concrete and reinforcement contributions, respectively; however, the parameters considered vary between the models. For example, the NZS 3101-06 (1995) and ASCE/SEI 43-05 models consider the influence of axial load on V_c (ACI 318 does not), the EC8-04 and ASCE/SEI 43-05 models include the impact of the vertical web reinforcement, and the detailed model of NZS 3101-06 (1995) uses $M/(Vl_w)$ instead of h_w/l_w, which is used by ACI 318-19, ASCE/SEI 43-05, and AIJ 1999.

The literature review by Rojas-León (2022) also includes an evaluation of wall shear strength equations reported in the literature, along with a description of the databases used in the calibration/validation of the models (Appendix A, Table A.2). For most of these studies, wall shear strength relations were developed by identifying relevant parameters based on a literature review, investigating the mechanics of the problem, and using statistical analysis of a data set or data sets. Subsequently, a calibration process was employed to fit the coefficients of the proposed model to the data; however, the performance of these equations was not typically checked against unseen data. Results presented in Table 1 enable a comparison of models analyzed in four studies (Sánchez-Alejandre and Alcocer 2010; Carrillo and Alcocer 2013; Kassem 2015; Looi and Su 2017) in terms of their mean and COV. As noted previously, the models are typically valid and perform well only when the parameters are within the ranges of the parameters used to calibrate the model. Because different databases were used and these databases used different criteria to determine which wall tests to include in the database, as well as different numbers of tests, different test parameters, and different ranges of test parameters, a model developed with a given database can be biased when it is evaluated with another database. Even if the ranges of relevant parameters are comparable, the size of the databases influences the reported means and COVs (Tanaka 1987). In addition, as databases become large, it often becomes infeasible to completely interpret the data using statistical models. In such cases, the application of ML is valuable (Dey 2016).

ML models

Although ML models can be powerful, they tend to be complex, challenging to use, and difficult to interpret (Bzdok et al. 2018). Several recent studies developed ML models to estimate wall shear strength (Chen et al. 2018; Moradi and Hariri-Ardebili 2019; Keshtegar et al. 2021; Feng et al. 2021); Appendix A, Table A.3 provides a summary of the databases, variables, and error indicators used in some of these studies. The results reported in Table A.3 demonstrate the potential and significant predictive power of ML models relative to other models (Table 1); however, these models suffer drawbacks, as described in the following paragraphs.

To train an ML model, a more extensive database is required; however, it also is critical to carefully screen the tests included in the database to ensure that they are aligned with the goals of the model being developed. For example, if the study is related to assessing wall shear strength, then the tests used in the database should include only walls that failed in shear, and outliers should be carefully reviewed to ensure that the data should be included (for example, inconsistent results are reported; an additional test variable is included that would impact results, such as corrosion; and the test variables satisfy code-minimum requirements, such as material properties). Model performance should be reported, including a well-known error indicator such as the mean and COV of the true-to-predicted ratio, to facilitate comparisons. Also, if an ML model is compared with other models (for example, Table 1), then the comparison should also include results of other (adequately trained) ML models to judge the performance of the ML model. ML models are more complex than models developed based on a (simple) equation; therefore, better performance is expected. If this is not the case, it implies that a complex model has similar (or worse) performance than a simple model; therefore, the added complexity is redundant because the simple model already captures the relevant patterns and relationships in the data.

A vast majority of structural tests reported in the literature were conducted at less than full-scale (for example, one-fifth to three-quarters); therefore, it is essential to develop models using dimensionless and/or mechanics-based normalized variables (for example, aspect ratio versus wall height and wall length, stress versus force) such that the database and model results are representative of both reduced-scale tests and full-scale components (for example, walls). The need for this step becomes clear when the relationship between the predicted variable and the error indicator selected for the optimization problem is evaluated. For example, if shear strength is the variable being estimated and an error such as the root-mean-squared error (RMSE) is used to train the

*The Appendix is available at **www.concrete.org/publications** in PDF format, appended to the online version of the published paper. It is also available in hard copy from ACI headquarters for a fee equal to the cost of reproduction plus handling at the time of the request.

Table 1—Wall shear strength model comparisons reported in different studies: V_{true}/V_{pred}

	Sánchez-Alejandre and Alcocer (2010)		Carillo and Alcocer (2013)		Kassem (2015)		Looi and Su (2017)	
Model	Mean	COV	Mean	COV	Mean	COV	Mean	COV
ACI 318-19, Ch. 18*	1.43	0.26	0.82	0.24	1.65	0.37	1.01	0.37
ACI 318-11, Ch. 11	—†	—	0.90	0.21	—	—	—	—
ACI 318-14, Ch. 11	—	—	—	—	—	—	0.96	0.37
AIJ (1999)	1.00	0.27	—	—	—	—	—	—
CSA A23.3-14	—	—	—	—	—	—	1.35	0.44
EC8 (2004)	—	—	—	—	2.54	0.71	—	—
Barda et al. (1977)	—	—	—	—	1.39	0.47	—	—
Wood (1990)	0.99	0.24	—	—	0.78	0.32	—	—
Hwang and Lee (2002)	1.06	0.22	—	—	1.26	0.56	—	—
Sánchez-Alejandre and Alcocer (2010)	1.00	0.13	0.79	0.12	1.91	0.29	0.84	0.35
Gulec and Whittaker (2011)	—	—	1.06	0.09	1.34	0.24	0.89	0.31
Carrillo and Alcocer (2013)	—	—	1.00	0.08	—	—	—	—
Kassem (2015)	—	—	—	—	1.00	0.21	—	—
Looi and Su (2017)	—	—	—	—	—	—	1.04	0.27

*Sánchez-Alejandre and Alcocer (2010) use ACI 318-08 Ch. 21; Carrillo and Alcocer (2013) and Kassem (2015) use ACI 318-11 Ch. 21; Looi and Su (2017) use ACI 318-14, Ch. 18. These equations are same as those in ACI 318-19 Ch.18.

†Model not included in comparison.

model, then the direct difference between the observed value (or true value, V_{true}) and the predicted shear strength value (V_{pred}) is minimized (that is, not a percentage error), which can lead to large errors for lower values of V_{true}. Another option is to use an error indicator equal to the difference in the true-to-predicted value; however, this does not address the issue of reduced scale tests. Finally, the coefficient of determination ($R^2 \in [0,1]$) is another error indicator that is commonly used; however, results can be misleading because this approach compares a given model to the null model, and larger R^2 values can be obtained for less precise models (Barret 1974).

For ML models, it is common to use two data sets: a training set and a testing set. The training set is used to train (calibrate) the model, and the testing set is used to verify that the trained model will perform similarly when predicting unseen data. Acceptable performance is achieved where the value of the error obtained for the testing set is comparable to that obtained for the training set. Although this comparison should be carefully addressed and ideally verified in terms of the error used in the optimization process and other meaningful error indicators to demonstrate model robustness, this added step is often not adequately considered.

FRAMEWORK

The proposed framework uses statistical and ML approaches to establish target performance requirements for component capacity models (for example, models for column and wall shear strength, beam flexural strength, reinforcement development length, and so on), or other models with similar characteristics, with different levels of complexity based on the use of a common, comprehensive database. The framework overcomes the limitations highlighted in the Introduction by allowing the user to define model performance requirements based on the desired level of model complexity.

The proposed framework adopts the generic steps of ML—that is, collection and preparation of data, feature selection, selection of ML algorithms, selection of model and hyper-parameters, model training, and model performance evaluation (Alzubi et al. 2018)—but also requires specific sub-steps to: a) define relevant (starting) features based on the mechanics of the problem; b) address the issue of using reduced-scale tests to predict capacities of full-scale specimens; c) develop an iterative sensitivity analysis to train the ML model; and d) train Elastic Net Models (ENMs) using engineered features defined from the starting features. Each of these steps is described in detail in the following subsections.

Step 1: Collection and preparation of data

A data set of walls with reported flexure-shear (F-S), diagonal-tension (D-T), or diagonal-compression (D-C) failure modes was obtained using the UCLA-RC Walls Database, which includes detailed and parametrized information on more than 1100 RC wall tests (Abdullah and Wallace 2018, 2021; Abdullah 2019). Tests with incomplete material test information were excluded because this information is required to define the variables used in this study. The reduced data set included a total of 412 wall tests. The dataset was further evaluated resulting in the removal of 72 tests because: a) test walls included artificial cracks to study corrosion (six tests, Zheng et al. [2015]); b) reported lateral load readings did not match the values reported in figures provided in various papers or reports (nine tests, Li and Li [2002]); c) test walls had asymmetric cross-sectional

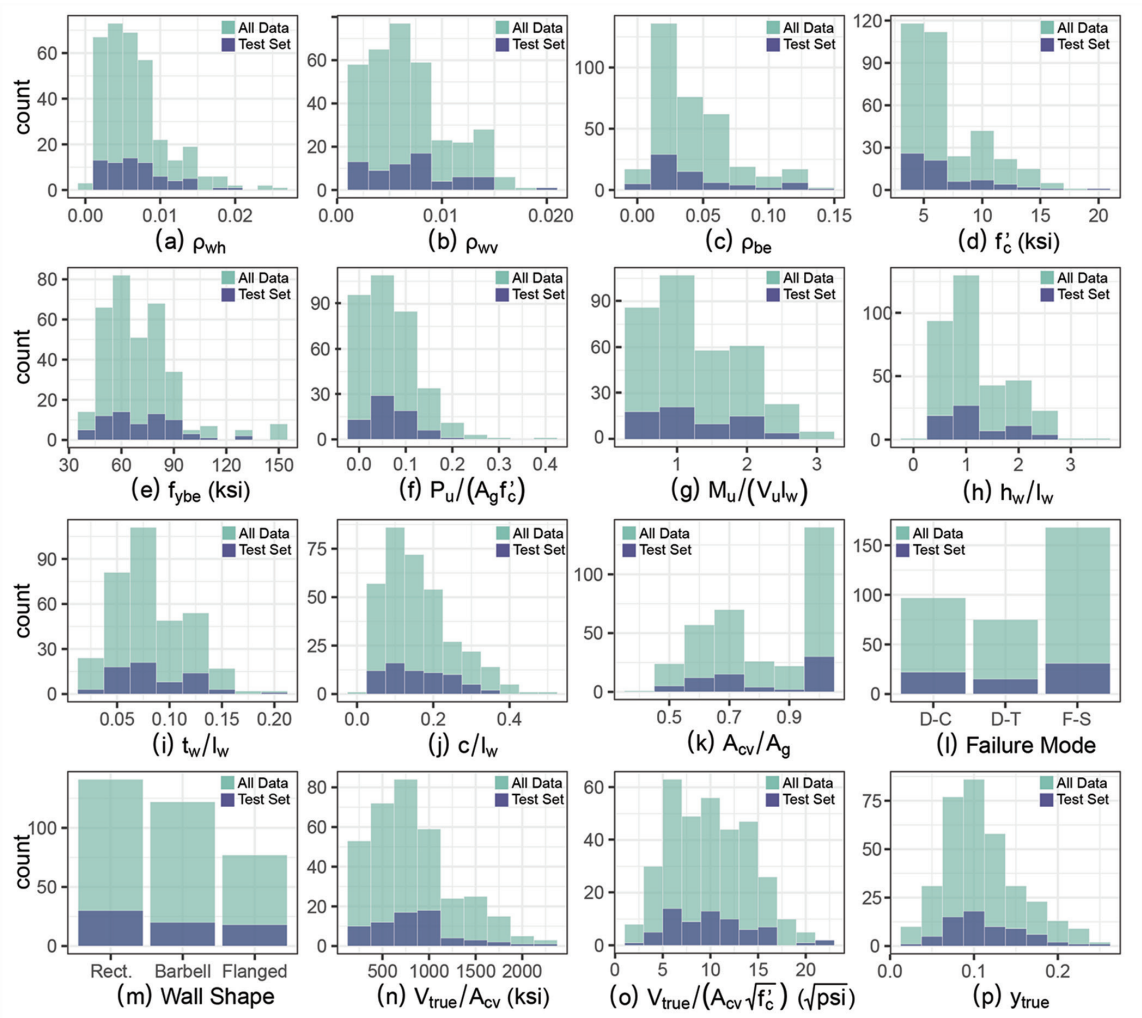

Fig. 1—Histograms of relevant parameters. (Note: 1 ksi = 6.895 MPa; 1 kip = 4.448 kN.)

shapes such as T-shape, L-shape, half barbell, and wing walls (20 tests); or d) reported values for tested compressive strength of concrete f_c' were less than the limit of 20.7 MPa (3.0 ksi) given in ACI 318-19 for special seismic systems (37 tests). Asymmetric walls were excluded because the number of walls with these cross-sectional shapes was low (20) compared to the number of rectangular, barbell, or flanged walls. Thus, if these tests are incorporated into the larger dataset of symmetric walls, the optimization process will likely overlook the inherent differences between asymmetric and symmetric walls. A more appropriate approach in this case, as implemented by Rojas-León (2022), is to develop a model excluding asymmetric wall cross sections and then evaluate whether simple changes to the model could be implemented to address shear strength estimates for the asymmetric walls.

Based on the aforementioned filters, a final (clean) dataset of 340 symmetric wall tests was obtained (refer to the Appendix) and randomly split into a training set with 80% of the tests (272) and a testing set with 20% of the tests (68) to verify the performance of the models. Figure 1 compares histograms for various database parameters of the entire data set and the testing set, where f_c' is the specified compressive strength of concrete; ρ_{be} is the boundary region longitudinal reinforcement ratio; f_{ybe} is the specified yield strength of the boundary region longitudinal reinforcement; ρ_{wh} and f_{ywh} are the ratio and specified yield strength of the horizontal web reinforcement, respectively; ρ_{wh} and f_{ywv} are the ratio and specified yield strength of the vertical web reinforcement, respectively; P_u, M_u, and V_u are the measured axial load, moment, and shear, respectively; l_w is the wall length in the direction of the applied shear force; h_w is the total wall height; A_{be} is the cross-sectional area bounding the longitudinal reinforcement at a wall boundary; A_{cv} is the cross-sectional area bounded by the wall length and the web thickness (t_w); A_g is the gross cross-sectional area; c is the neutral axis depth; and y_{true} is the normalized shear stress (introduced later).

Step 2: Defining ML models and features

This step involves identifying the potentially relevant parameters based on a literature review and studying relatively simple mechanics-based models and appropriate free-body diagrams. For this application—that is, RC wall shear strength—a free-body diagram of a wall with a diagonal crack was used. Based on these considerations, the following relationships were derived (Rojas-León 2022).

$$V_u \propto A_g f_c' \tag{1}$$

Table 2—ENMs definition

Model	Short reference	Long reference
ENM1	$y \sim \mathbf{X}$	$y_j \sim N(\mu_j, \sigma),\ \mu_j = \mathbf{x}_j^T\beta,\quad \forall j \in \{1, 2, \ldots, n\}$
ENM2	$y \sim \tilde{\mathbf{X}}$	$y_j \sim N(\mu_j, \sigma),\ \mu_j = \tilde{\mathbf{x}}_j^T\beta,\quad \forall j \in \{1, 2, \ldots, n\}$
ENM3	$\sqrt[3]{y} \sim \tilde{\mathbf{X}}$	$\sqrt[3]{y_j} \sim N(\mu_j, \sigma),\ \mu_j = \tilde{\mathbf{x}}_j^T\beta,\quad \forall j \in \{1, 2, \ldots, n\}$
ENM4	$\log(y) \sim \tilde{\mathbf{X}}$	$\log(y_j) \sim N(\mu_j, \sigma),\ \mu_j = \tilde{\mathbf{x}}_j^T\beta,\quad \forall j \in \{1, 2, \ldots, n\}$
ENM5	$y \sim \mathbf{X}_{\boldsymbol{poly}}$	$y_j \sim N(\mu_j, \sigma),\ \mu_j = \mathbf{x}_{poly_j}^T\beta,\ \forall j \in \{1, 2, \ldots, n\}$
ENM6	$\sqrt[3]{y} \sim \mathbf{X}_{\boldsymbol{poly}}$	$\sqrt[3]{y_j} \sim N(\mu_j, \sigma),\ \mu_j = \mathbf{x}_{poly_j}^T\beta,\ \forall j \in \{1, 2, \ldots, n\}$
ENM7	$\log(y) \sim \mathbf{X}_{\boldsymbol{poly}}$	$\log(y_j) \sim N(\mu_j, \sigma),\ \mu_j = \mathbf{x}_{poly_j}^T\beta,\ \forall j \in \{1, 2, \ldots, n\}$
ENM8	$y \sim \tilde{\mathbf{X}}_{\boldsymbol{poly}}$	$y_j \sim N(\mu_j, \sigma),\ \mu_j = \tilde{\mathbf{x}}_{poly_j}\beta,\ \forall j \in \{1, 2, \ldots, n\}$
ENM9	$\sqrt[3]{y} \sim \tilde{\mathbf{X}}_{\boldsymbol{poly}}$	$\sqrt[3]{y_j} \sim N(\mu_j, \sigma),\ \mu_j = \tilde{\mathbf{x}}_{poly_j}\beta,\ \forall j \in \{1, 2, \ldots, n\}$
ENM10	$\log(y) \sim \tilde{\mathbf{X}}_{\boldsymbol{poly}}$	$\log(y_j) \sim N(\mu_j, \sigma),\ \mu_j = \tilde{\mathbf{x}}_{poly_j}^T\beta,\ \forall j \in \{1, 2, \ldots, n\}$

Note: n is number of features model uses.

$$V_u \overset{\propto}{\sim} \rho_{wh} f_{ywh} h_w t_w \tag{2}$$

$$V_u h_{eff} \overset{\propto}{\sim} \rho_{wv} f_{ywv} (l_w - c)^2 t_w \tag{3}$$

$$V_u h_{eff} \overset{\propto}{\sim} \rho_{be} f_{ybe} A_{be} l_w \tag{4}$$

$$V_u h_{eff} \overset{\propto}{\sim} \left(f_c' + \frac{P}{A_g}\right) t_w c^2 \tag{5}$$

$$V_u h_{eff} \overset{\propto}{\sim} \left(f_c' - \frac{P}{A_g}\right) t_w (l_w - c)^2 \tag{6}$$

Based on these relationships, and to address the use of reduced-scale test specimens, the following 10 non-dimensional variables are selected and are named the "starting features." These variables can be identified by normalizing V_u by $A_g f_c'$ in Eq. (1) through (6) and by applying reasonable approximations in some cases (for example, considering c as a fraction of l_w, and neglecting constants because the model calibration process will address this).

$$x_1 = \rho_{wh}(f_{ywh}/f_c') \tag{7}$$

$$x_2 = \rho_{wv}(f_{ywv}/f_c') \tag{8}$$

$$x_3 = \rho_{wbe}(f_{ybe}/f_c') \tag{9}$$

$$x_4 = 1 + P_u/(A_g f_c') \tag{10}$$

$$x_5 = c/l_w \tag{11}$$

$$x_6 = M_u/(V_u l_w) \tag{12}$$

$$x_7 = t_w/l_w \tag{13}$$

$$x_8 = t_w/h_w \tag{14}$$

$$x_9 = h_w/l_w \tag{15}$$

$$x_{10} = A_{be}/A_g \tag{16}$$

The predicted variable is the normalized shear strength defined as

$$y_{true} = V_{true}/(A_g f_c') \tag{17}$$

Between $(1 + P_u/(A_g f_c'))$ and $(1 - P_u/(A_g f_c'))$, only one option is considered because they are related to the same parameters in Eq. (5) and (6), and because the presence of a constant (that is, "intercept" or equivalent) in the calibration process would suggest dropping one of the terms because it is linearly dependent on the other. The height of the wall used to define the effective flange width according to ACI 318-19 Section 18.10.5.2 was estimated as $h_w \approx$ effective height (h_{eff})/0.7, where h_{eff} corresponds to the shear span, defined as M_u/V_u. It is well established that flanged walls have a larger shear strength (Gulec et al. 2009; Gulec and Whittaker 2011; Kassem 2015; Kim and Park 2020); thus, cross-sectional area A_g is used instead of A_{cv} in Eq. (17).

The feature matrix $\mathbf{X}$ contains the 10 starting features defined in Eq. (8) through Eq. (17). The following feature matrixes ($\tilde{\mathbf{X}}$, $\mathbf{X}_{\boldsymbol{poly}}$, and $\tilde{\mathbf{X}}_{\boldsymbol{poly}}$) are obtained by using feature engineering. Feature matrix $\tilde{\mathbf{X}}$ contains 140 features because the following 14 functions were applied to the 10 original (starting) features: identity function, $(\cdot)^{-1}$, $(\cdot)^2$, $(\cdot)^{-2}$, $(\cdot)^{1/2}$, $(\cdot)^{-1/2}$, $(\cdot)^3$, $(\cdot)^{-3}$, $(\cdot)^{1/3}$, $(\cdot)^{-1/3}$, $exp(\cdot)$, $exp(-\cdot)$, $\log(\cdot)$, and $-\log(1 + \cdot)$. Feature matrix $\mathbf{X}_{\boldsymbol{poly}}$ has 285 features (combining the 10 starting features with a cubic polynomial). Feature matrix $\tilde{\mathbf{X}}_{\boldsymbol{poly}}$ has 679 features that are obtained by combining the 14 more significant features of the $\tilde{\mathbf{X}}$ matrix with cubic polynomial coefficients. Cubic polynomials were used because Eq. (1) through (6) can be formed by multiplying up to three starting features. Also, to reduce skewness or highlight trends, other variations of the output variable y (refer to Eq. (17)) are defined as $\sqrt[3]{y}$ and $\log(y)$. The subset of 14 more significant features of $\tilde{\mathbf{X}}$ is obtained after performing the sensitivity analysis (explained later) for ENM2 (introduced later in Table 2).

The starting features will be the input parameters of one or more complex ML models, which will predict the normalized shear stress defined in Eq. (17). The selected complex ML models for this study are the artificial neural network (ANN) and Random Forest (RF) regression models because they are applicable for this study (the predicted parameter is a continuous variable), and because they are well-known models that are not complicated to implement in programming languages (for example, Matlab, R, and Python, which have various built-in functions to simplify their implementation). The starting and engineered features are also used to create a suite of ENMs; a total of 10 ENMs are defined (refer to Table 2).

ENMs (Zou and Hastie 2005) are a simple and more interpretable ML model type because they are a penalized linear modeling approach with a mixture of ridge regression (Hoerl and Kennard 1970) and Least Absolute Shrinkage and Selection Operator (LASSO) regression (Tibshirani 1996). Ridge regression reduces the impact of collinearity on the features, whereas LASSO reduces the dimension of the problem by shrinking some of the coefficients to zero (less significant

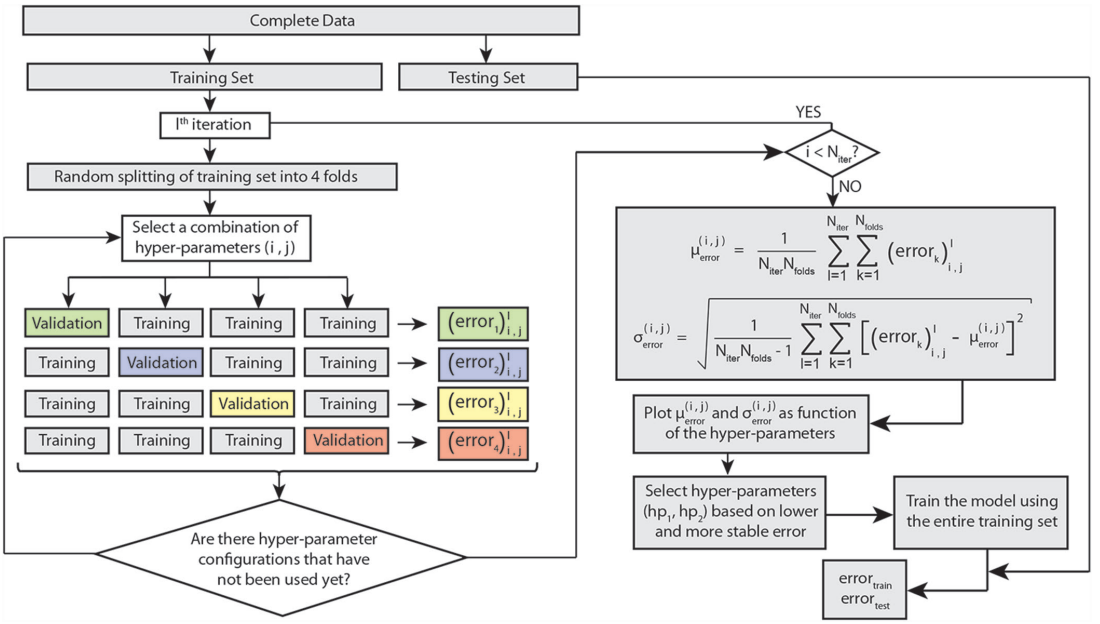

Fig. 2—Sensitivity analysis algorithm to select optimum set of hyper-parameter values.

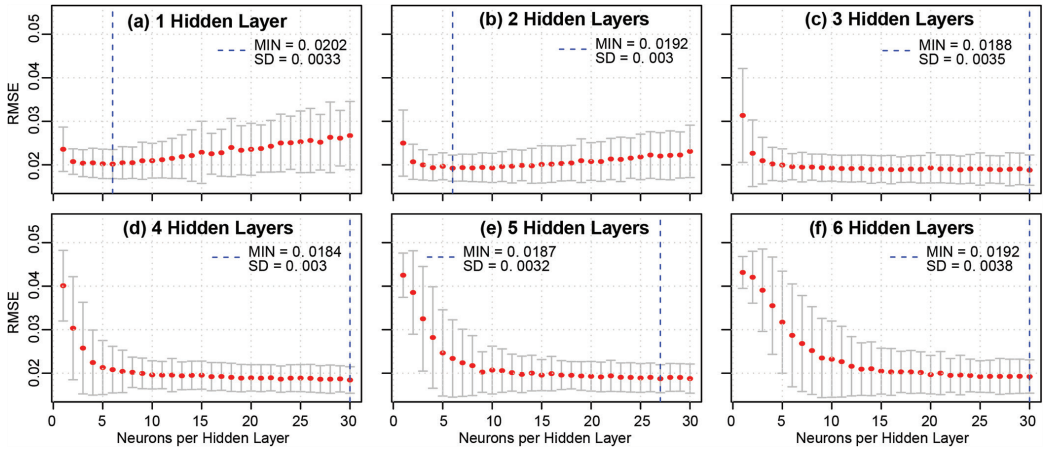

Fig. 3—Sensitivity analysis results to define optimum ANN configuration.

parameters). ENMs have two hyper-parameters: a) $\lambda > 0$ is the complexity parameter that controls the weight of the penalization factors; and b) $\alpha \in [0,1]$ is the compromise between Ridge ($\alpha = 0$) and LASSO ($\alpha = 1$). Small λ values can result in an overfitted model (too complex), whereas high λ values can result in an underfitted model (too simple).

Step 3: Sensitivity analysis and selection of hyper-parameters

The hyper-parameter sensitivity analysis described in Fig. 2 was implemented for the 12 ML models (1 ANN, 1 RF regression, and 10 ENMs) using an iterative k-fold cross-validation (CV) method with, in this case, $N_{iter} = 100$ iterations and $k = 4$ folds. K-fold CV is useful for data scientists when dealing with small databases (a few thousand data samples). Iterations are included because, in Structural Engineering, the database is typically even smaller (just a few tens or hundreds). The number of folds was set as $k = 4$ because it makes the validation set representative of the testing set (that is, the same size). Once the sensitivity analysis is completed, $k \times N_{iter} = 4 \times 100 = 400$ RMSE values are computed for each configuration of hyper-parameters. The mean and standard deviation of the calculated RMSEs are obtained for each of these configurations. The optimal hyper-parameters are selected based on the lower mean error and standard deviation.

ANN—Rules suggesting values for the number of hidden layers and neurons per layer (main hyper-parameters) can be found in the literature (Chen et al. 2018; Moradi and Hariri-Ardebili 2019), which are covered by the ranges selected for the sensitivity analysis. The sensitivity analysis results are shown in Fig. 3; a blue dashed line indicates the best ANN configuration for each number of hidden layers considered. From there, the optimum ANN is the one with four hidden layers and 30 neurons in each layer because it has the lowest mean error (MIN RMSE) and the lowest standard deviation (SD). Previous configurations (the same four hidden layers, but fewer neurons) show an extensive range of similar and stable results.

RF regression—A large number of decision trees (1000 trees) are selected to ensure that a stable error level is reached. For this study, the error became stable at approximately 300 trees. Two other hyper-parameters could have

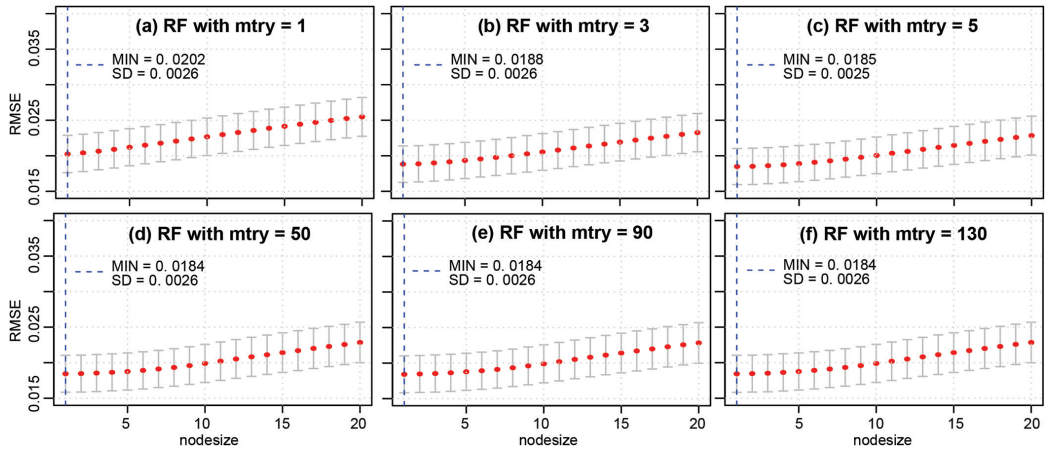

Fig. 4—Sensitivity analysis results to define optimum RF configuration.

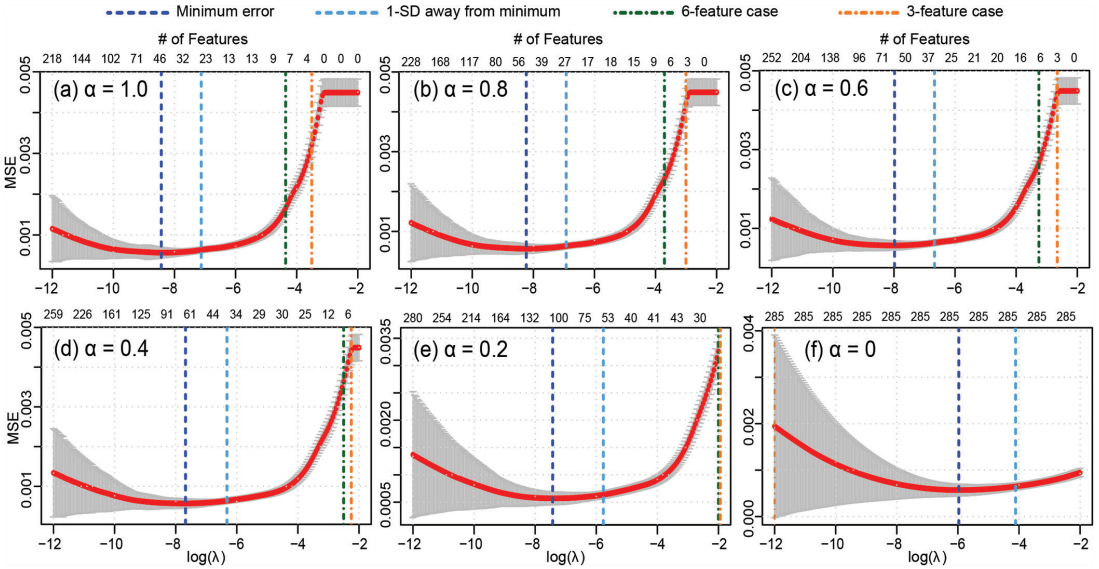

Fig. 5—Sensitivity analysis results for ENM6 model: $\sqrt[3]{y_j} \sim N(\mu_j, \sigma), \mu_j = x_{poly_j}'\beta$.

an impact on the performance of the model (Zhang and Ma 2012): 1) the number of variables (selected among all the features) in each cell (*mtry*); and 2) the prespecified threshold of maximum observation per cell (*nodesize*). The sensitivity analysis covered ranges of values for both hyper-parameters according to observations by Zhang and Ma (2012) concerning *mtry*, and according to Breiman (2001) and Segal and Xiao (2011) concerning *nodesize*. Figure 4 shows that the RF results are only slightly sensitive to *mtry*, and that having large trees (small *nodesize*) results in low errors (RMSE). Optimal hyper-parameters are selected as *mtry* = 50 and *nodesize* = 1 (minimum mean and SD).

ENMs—The log(λ) values ranged from –12 to –2, while the α values were 1.0, 0.8, 0.6, 0.4, 0.2, and 0.0 for all 10 ENMs defined in Table 2. As an example, Fig. 5 presents the sensitivity analysis results for ENM6; the optimum version of the model is indicated with a black dashed vertical line, while the blue, green, and orange vertical dashed lines indicate the λ values associated with the selected underfitting levels. In this study, three levels of underfitting are selected: 1) one in which the error corresponds to the error that is one standard deviation away from the error of the optimum model ("1-SD away" version); 2) an underfitted model that uses six features only ("6-feature" version); and 3) an underfitted model that uses three features only ("3-feature" version). These underfitting levels were selected because they are representative of models adopted in building codes and standards, and a model with this complexity level is of particular interest in this study.

Figure 6 presents the mean errors obtained from the sensitivity analysis of the 10 ENM models and demonstrates that regardless of the value of α considered, there is a λ value where practically the same optimum error is reached. Figure 7 indicates that it is difficult for the models to exclude features to achieve the defined underfitted levels of interest for lower α values. Because of these reasons, for each ENM in this study, the selected hyper-parameter configuration for the optimum and underfitted complexity levels is α = 1, and its associated corresponding λ value—that is, all selected ENMs are LASSO models.

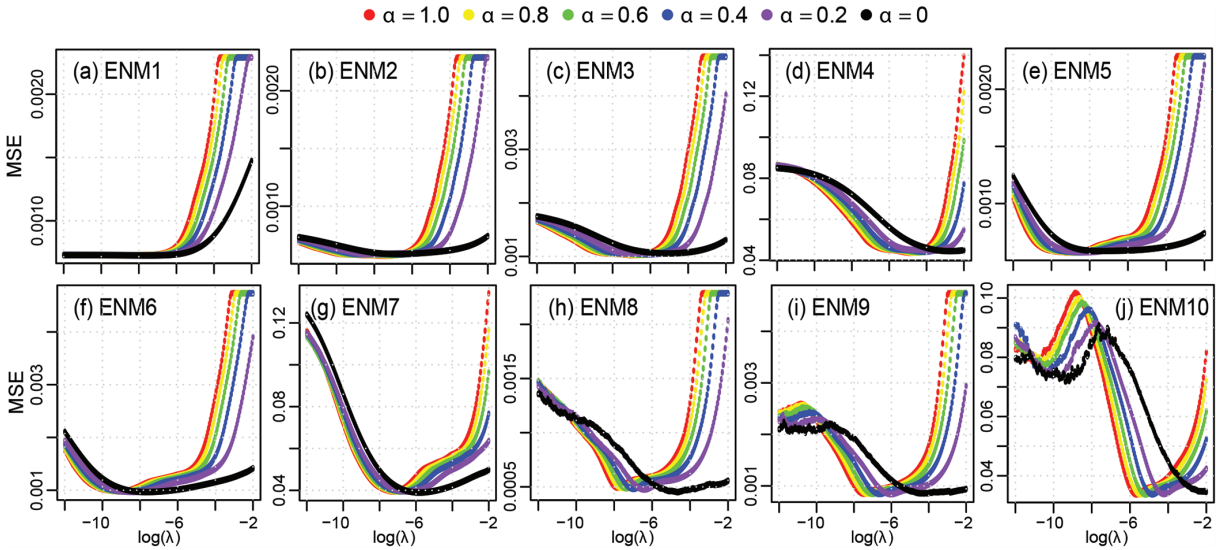

Fig. 6—Mean of MSE obtained from sensitivity analysis of all ENM models.

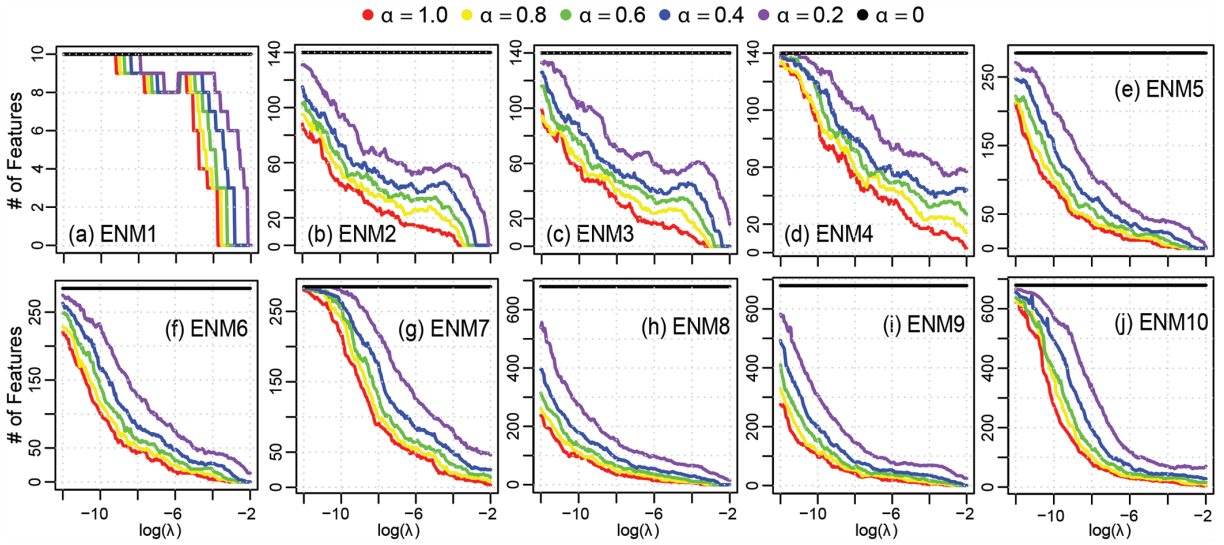

Fig. 7—Mean non-zero coefficients obtained from sensitivity analysis of all ENM models.

Step 4: Training, verification, and selection of best-performing models

All the models are trained using the training set with selected sets of hyper-parameters. This results in 42 trained ML models: the optimum ANN (1 model), the optimum RF Regression (1 model), the optimum version of each LASSO (10 models), the 1-SD away version of each LASSO (10 models), the 6-feature version of each LASSO (10 models), and the 3-feature version of each LASSO (10 models). The acceptability criterion adopted in this study defines a model as acceptable when the errors of the training and testing sets are both within a defined margin away from the converging error, which is ±20% for the optimum models and ±10% for the underfitted models. The converging error is taken as the average of the training and testing errors. Optimum models have a larger margin because they are right on the balanced point between the underfitted and overfitted models. Thus, they have the potential to "keep learning" (for example, re-adjust their coefficients a bit) if new data are provided for training. On the other hand, by definition, underfitted models are not capable of capturing enough details, and they follow more rough trends identified from the data, which is the reason for the stricter margin around the converging error. The performance of existing models and the training of several new models also informed the selection of the acceptable bandwidth around the converging error (Rojas-León 2022).

Optimum ANN and RF—Although the training process was based on the RMSE between y_{true} and y_{pred} (refer to Fig. 8(a) and 9(a)), similar model performance (that is, training and testing errors within ±20% of the converging error) is verified when using the predicted values from the training and testing sets to compute the mean and COV of V_{true}/V_{pred} for the optimum ANN (Fig. 8(b) and (c)) and optimum RF (Fig. 9(b) and (c)). Figures 8(c) and 9(c) also show that the predictions for the training and testing sets have the same distribution shapes.

Optimum and underfitted LASSO models—All 40 LASSO models selected (four from each ENM defined in Table 2) are trained using only the features associated with each chosen hyper-parameter configuration—that is, they are linear

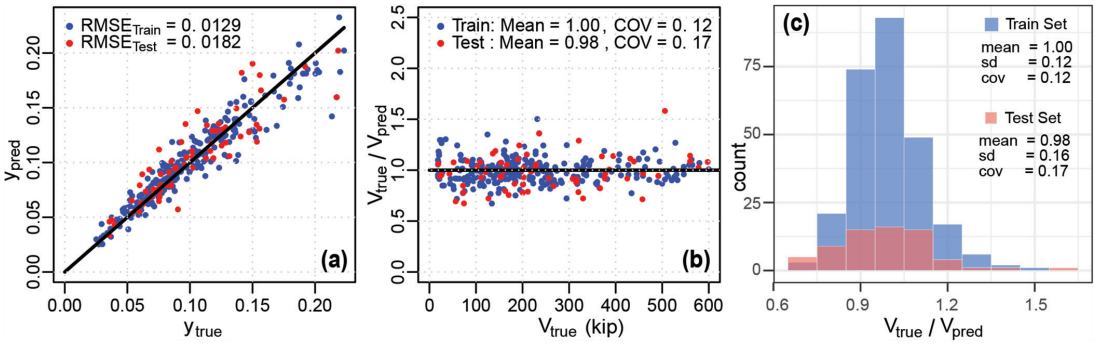

Fig. 8—Performance of selected ANN on training and testing sets in terms of: (a) normalized shear stress; (b) V_{true}/V_{pred} *versus* V_{true}*; and (c) distributions of* V_{true}/V_{pred}*. (Note: 1 kip = 4.448 kN.)*

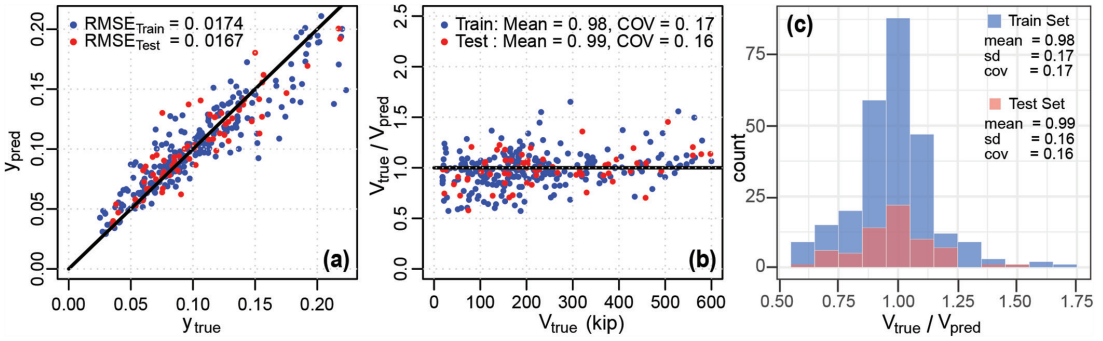

Fig. 9—Performance of selected RF on training and testing sets in terms of: (a) normalized shear stress; (b) V_{true}/V_{pred} *versus* V_{true}*; and (c) distributions of* V_{true}/V_{pred}*. (Note: 1 kip = 4.448 kN.)*

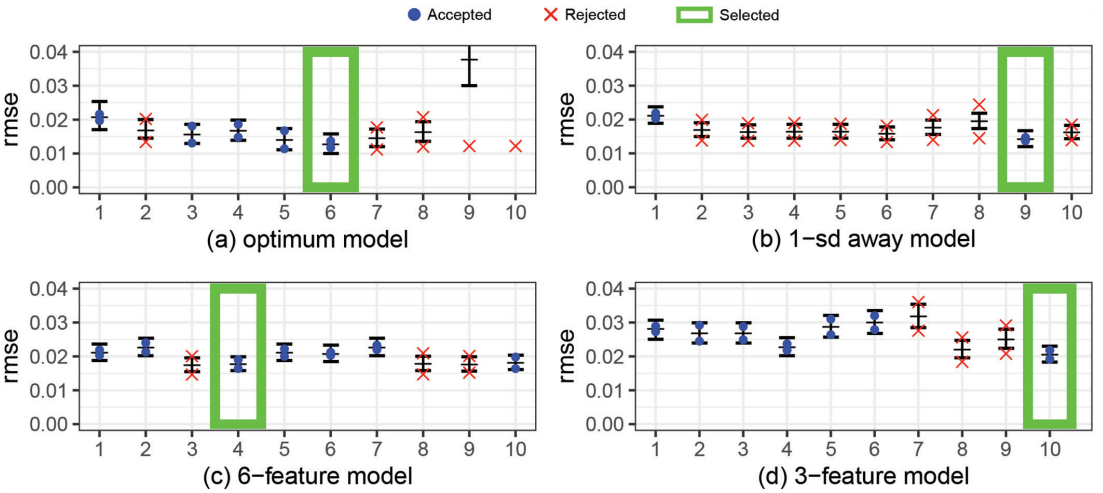

Fig. 10—Selection of best LASSO model for each complexity level based on acceptability criterion.

regressions with different engineered features. Figures 10(a) to (d) show the training and testing errors (RMSE between y_{true} and y_{pred}) with a blue dot or a red cross, depending on whether they meet or do not meet the acceptability criterion, respectively. For each complexity level, the model meeting the acceptability criterion with the smaller converging error was selected (which are highlighted with a green box in Fig. 10; full-color version can be accessed at www.concrete.org); the best optimal LASSO model, the best 1-SD away LASSO model, the best 6-feature LASSO model, and the best 3-feature version. Note that the optimum LASSO models No. 9 and 10 have significant errors, which is attributed to the implemented automated selection of hyper-parameters that are just a little past the underfitted-overfitted sweet spot, which is the reason that the 1-SD away model was included (especially for those LASSO models that are more complex). Figure 11 verifies the good and similar performance (training versus testing errors) in terms of the same error indicators used for the optimum ANN and optimum RF. The error goes up gradually, and distributions of the y_{true}/y_{pred} become wider as the complexity level of the models is relaxed. Nonetheless, the errors obtained for the 6-feature and 3-feature linear regressions are still very low compared to the results of existing equations in Table 1.

Except for the RF regression, all the learning curves shown in Fig. 12 (also obtained by iterating at each set size)

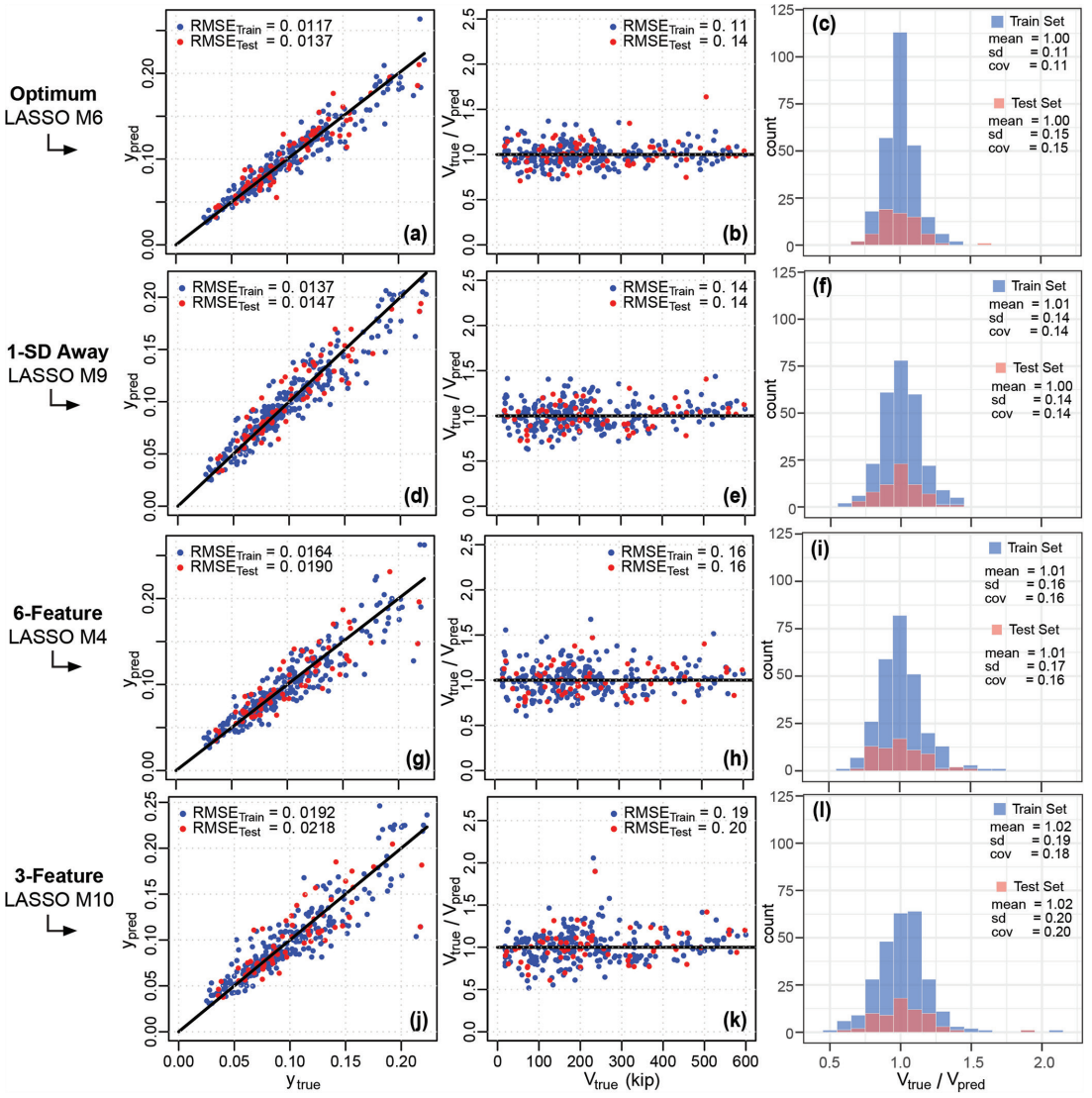

Fig. 11—Performance of selected LASSO models on training and testing sets in terms of: (a), (d), (g), and (j) normalized shear stress; (b), (e), (h), and (k) V_{true}/V_{pred} versus V_{true}; and (c), (f), (i), and (l) distributions of V_{true}/V_{pred}. (Note: 1 kip = 4.448 kN.)

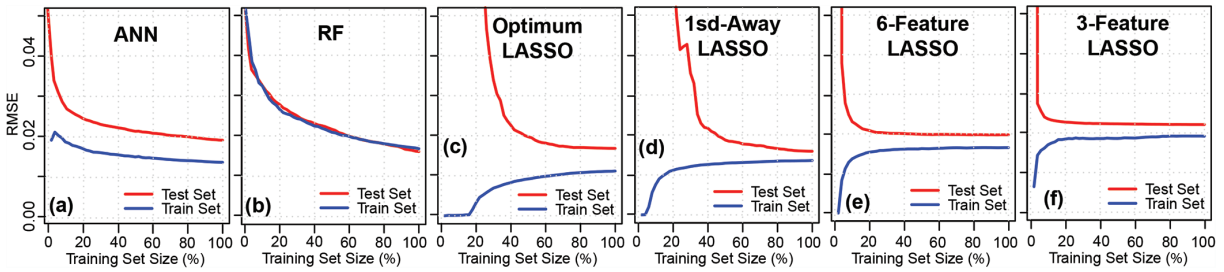

Fig. 12—Learning curves of selected and verified ML models.

have a gap between the training and testing curves and reach a plateau when approaching the use of 100% of the training set. Because of this, the error obtained when including future data into the training set to refine these same models (that is, keeping the same hyper-parameters and relevant features already identified) should fall between the training and testing errors, but closer to the training error. On the other hand, the training and testing learning curves for RF regression are very close to each other because a very large number of trees are selected. However, it is observed that the slope of the learning curves reduces (reaches a plateau) when the training size becomes larger. This behavior means that, if additional data are provided for training the same RF regression, the converging error would get closer to that plateau, resulting in a slightly lower error.

Step 5: Setting target errors for different model complexity levels

Because the six selected models demonstrate good performance that has been verified, adding data with a distribution

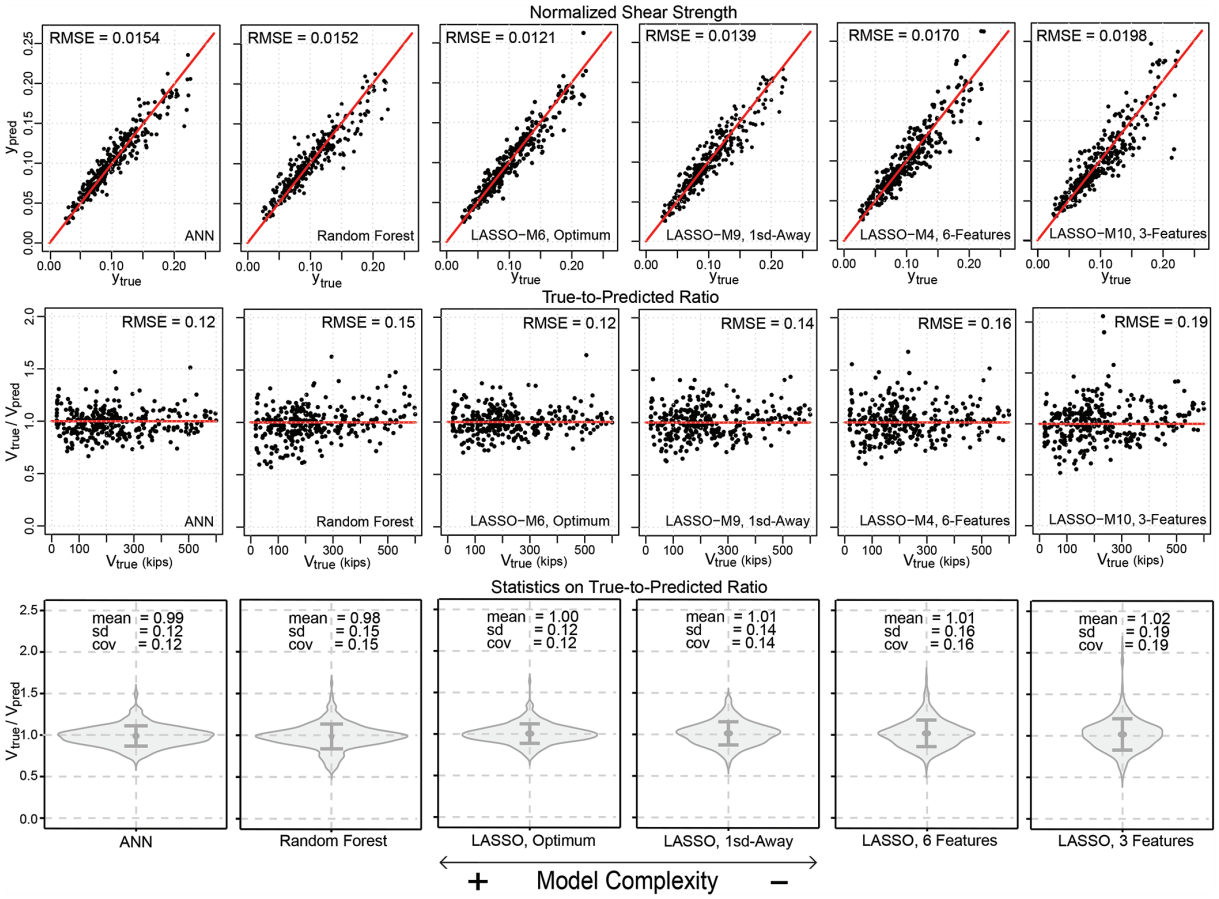

Fig. 13—Selected and trained ML models applied to entire database.

similar to that of the training data will only refine these models (Amazon Web Services 2019). Thus, the testing set is incorporated into the database used to train these models, but the same selected hyper-parameters and input features are kept. The results are presented side by side in Fig. 13, sorted from higher model complexity level (left) to lower model complexity level (right): optimum ANN, optimum RF, optimum LASSO M6, 1-SD away LASSO M9, 6-feature LASSO M4, 3-feature LASSO M10. As expected, performances are similar to those obtained previously and aligned with the observations derived from the learning curves. This good behavior is verified for all different error indicators used before.

Figure 13 shows that the ANN performs better (smaller error) than RF, but there is still room for the RF to improve if additional data are added to the database. The optimum LASSO model performs practically the same as the optimum ANN, or even slightly better if the RMSE between y_{true} and y_{pred} is considered. This is a relevant finding for two primary reasons: 1) the LASSO model is much less complex than the ANN model because, as noted before, LASSO models are linear regressions using those selected features only (which, for the optimum LASSO model, are 45 features engineered from the 10 starting features); and 2) the underfitted LASSO models can be understood as a smooth relaxation away from the optimum when looking for target model performances (errors) that fulfill user requirements for less complex models. The 1-SD away LASSO model is a linear regression of 14 features engineered from seven of the 10 starting features (x_1, x_2, x_3, x_6, x_8, x_9, and x_{10}), the 6-feature LASSO model is a linear regression of six features engineered from six of the 10 starting features (x_1, x_2, x_3, x_4, x_6, and x_8), and the 3-feature LASSO model is a linear regression that uses three features engineered from five of the 10 starting features (x_1, x_2, x_3, x_6, and x_{10}). Unlike the ANN or RF regression models, the LASSO models could be easily implemented in an Excel spreadsheet. Therefore, for the comprehensive database used in this study or for a similar one (similar parameter ranges and distributions, as is the case of the testing set with respect to the entire database accordingly with Fig. 1), models with different levels of complexity noted should comply with the requirements stipulated in Table 3.

Table 3—Target performances for different model complexity levels

Requirements	Model complexity level	
	Complex ML models	Simplified models
Number of parameters	—	~3 to 6
V_{true}/V_{pred} mean ratio	0.99 to 1.01	0.98 to 1.02
COV	≤0.12	0.16 to 0.19
Training versus testing error margin	±20% of converging error	±10% of converging error

COMMENTS ON RELEVANT PARAMETERS

Among the starting features defined from Eq. (7) to (16), the ones used in the 6- and the 3-feature LASSO models defining the performance requirement for a

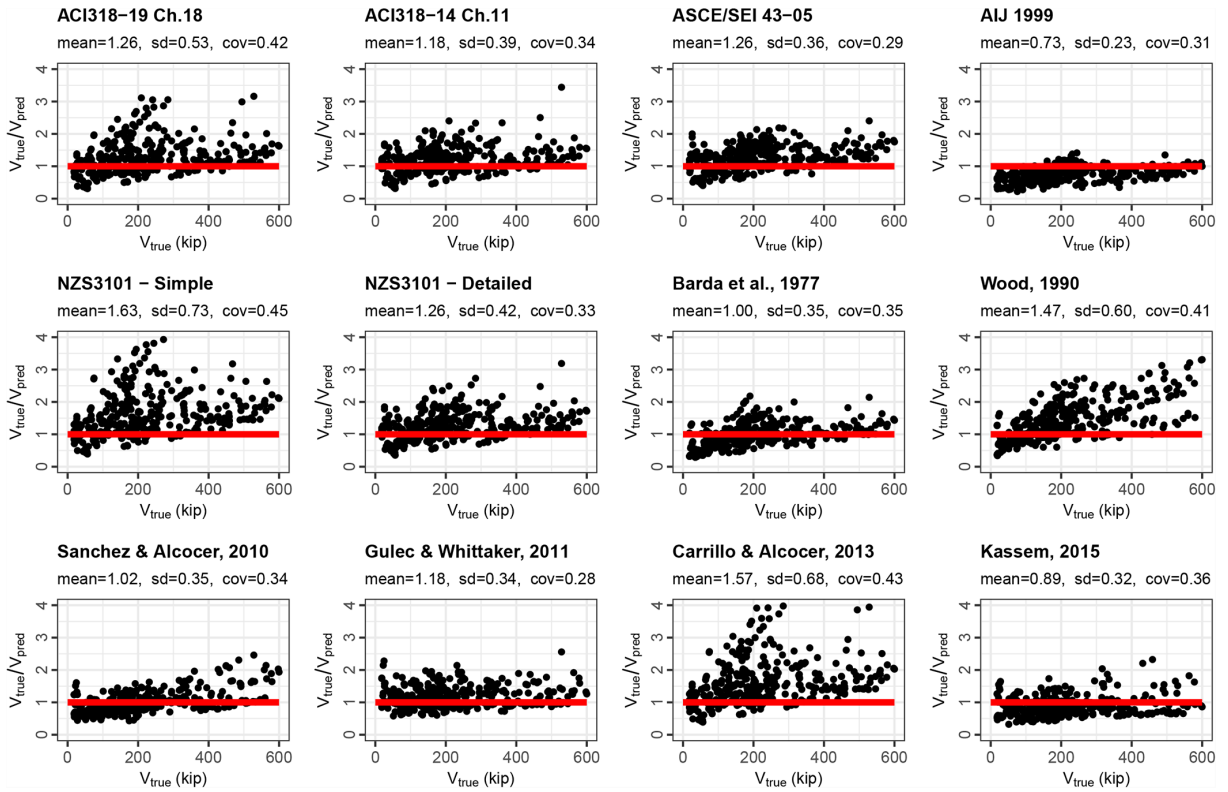

Fig. 14—Performance comparison of existing models using single, comprehensive database.

code-oriented equation are: $x_1 = \rho_{wh}(f_{ywh}/f_c')$, $x_2 = \rho_{wv}(f_{ywv}/f_c')$, $x_3 = \rho_{wbe}(f_{ybe}/f_c')$; $x_4 = 1 + P_u/(A_g f_c')$, $x_6 = M_u/(V_u l_w)$, $x_8 = t_w/h_w$, and $x_{10} = A_{be}/A_{cv}$. The only ones that are not listed herein are $x_5 = c/l_w$, $x_7 = t_w/l_w$, and $x_9 = h_w/l_w$. The absence of $x_5 = c/l_w$ can be attributed to the presence of $x_{10} = A_{be}/A_{cv}$ and $x_3 = \rho_{wbe}(f_{ybe}/f_c')$ because these two features can be used to represent the forces of compression or tension that are developed at the wall edges (and thus the neutral axis depth). The absence of $x_7 = t_w/l_w$ can be attributed to the presence of $x_8 = t_w/h_w$, which already accounts for the wall thickness and can be used together with $x_6 = M_u/(V_u l_w)$ to reproduce values that have a high correlation with $x_7 = t_w/l_w$. On the other hand, the absence of $x_9 = h_w/l_w$ might be surprising because some existing models used by codes or standards, or reported in the literature, use the wall aspect ratio (h_w/l_w) as a parameter to estimate wall shear strength (ACI 318-19 is one of those), whereas some other models use moment-to-shear span ratio ($M_u/(V_u l_w)$). In many of the tests reported in the literature (82% of the tests in the database used), these values are the same because the test involves a cantilever wall, fixed at the base, with a single point load applied near the top of the wall (that is, $M_u = V_u h_w$). For some tests reported in the literature, these values are not the same (for example, for a partial-height wall with an applied lateral load and moment at the top of the wall), and it is necessary to define an effective wall height $h_{w,eff}$ and wall aspect ratio ($h_{w,eff}/l_w$). As for the database used in this study, there are 32 specimens with a moment applied to the top of a partial height wall, three cantilever walls with two or more lateral loads, and 26 specimens tested with a double curvature configuration. For tests with multiple applied lateral loads (Cardenas and Magura 1972) or a moment applied at the top of the wall test (Segura and Wallace 2018), or both, the effective wall aspect ratio $h_{w,eff}$ was defined as M_u/V_u at the wall critical section (wall-foundation interface). If this approach is used, then identical results are produced from the wall test database using either h_w/l_w or $M_u/(V_u l_w)$. Thus, the constructed wall height was used to define $x_9 = h_w/l_w$ because that is how the aspect ratio has been defined in other studies. However, for the reasons given previously, it was expected that $x_6 = M_u/V_u l_w$ would be a better parameter to assess the shear strength of walls in buildings.

COMMENTS ON PERFORMANCES OF EXISTING MODELS

The performance of the existing models in codes and standards was evaluated using the common, comprehensive database gathered for this study (refer to Fig. 14). Upper limits (for example, $10\sqrt{f_c'}A_{cv}$ from the ACI 318-19 equation) were not considered to avoid introducing bias (conservatism) into the equations. Mean values varied from 0.73 to 1.63, and the COV values ranged from 0.28 to 0.45, and none of the existing models performed particularly well. The Gulec and Whittaker (2011) model had the least variation but with a mean value of 1.19. The Barda et al. (1977) model and the Sánchez-Alejandre and Alcocer (2010) model had mean values very close to 1.0, but COV values greater than 0.3. The ASCE 43-05 model (which is based on Barda et al. [1977]) resulted in a mean value of 1.26 and a COV of 0.29. None of these models satisfy the simplified model complexity level requirements stated in Table 3.

Also, none of the ML models analyzed in the literature review meet the target performance requirements for a complex model because the error is not small enough or

because the difference between the training and testing error is too large. The proposed framework provides a means to better train ML models, particularly when the addressed problem is based on basic mechanical principles. Applying the proposed framework resulted in similar results for all ML model types studied at their respective optimum complexity level. The observation that essentially the same performance of complex ML models (ANN, RF) was achieved with a simple LASSO model in this study indicates that the size of databases used for many civil engineering problems may still be too small to benefit from the use of complex ML model types because a linear regression with the right features has similar performance, or even slightly better. This is aligned with the rule of thumb that says ML models should be trained on at least an order of magnitude more samples than input model parameters (Morgan and Bourlard 1989; Google Developers 2022; Gonfalonieri 2019).

SUMMARY AND CONCLUSIONS

This study proposes a framework to obtain different target model performance requirements for models with different complexity levels. The approach is particularly useful when addressing a mechanics-based problem with a small database. The framework leads to properly trained machine-learning (ML) models (more than one) that enable the quantification of the gap between the performance of existing models and the best performance that can be achieved with currently available data; this allows the user to make informed decisions on the value of developing improved models (with less complexity than the ML models).

The framework is demonstrated by addressing the problem of assessing wall shear strength using a comprehensive database of 340 walls reported to have failed in shear. This application highlights how the framework can be used to address issues such as: a) existing reinforced concrete (RC) wall shear strength equations (used in building codes or standards or proposed in the literature) perform very differently when evaluated with different databases and the performance is generally poor when evaluated against a common, comprehensive database (that is, high error, high variance, or both); b) existing ML models were trained without addressing the issue that most databases are comprised of tests conducted at less than full scale or do not represent the spectrum parameters for as-built walls in buildings; and c) existing models with higher complexity suggest good performance by showing that they are better than models with less complexity, which is an unfair comparison. Finally, where models of similar complexity are compared, it is insufficient to conclude that the model with best performance should be selected because a third model with equivalent complexity could perform better—that is, model performance requirements are needed to guide this assessment.

When applied to the problem of assessing RC wall shear strength, the framework shows that a systematic methodology that recognizes the mechanics of the problem, the availability of limited data (compared to those databases with thousands or millions of samples available in fields where ML shows its greats potential), and avoids training issues such as those highlighted in this paper, can produce simple models with performance as good as (or nearly as good as) complex ML models. Because all ML models considered in this study at their optimum complexity level (ANN, RF regression, and LASSO model) result in very similar predictive performance, underfitted models derived from the optimum LASSO model can be taken as a smooth relaxation away from the optimum when looking for target model performance (errors) that fulfill user requirements for less-complex models.

For the application and database used in this study, the framework establishes a V_{true}/V_{pred} mean ratio very close to 1.0 with a COV in the range of 0.16 to 0.19 as the performance requirements for a less-complex model that could be used in codes and standards to predict RC wall shear strength. In addition, the training and testing errors should be within a margin of ±10% of the converging error (at least, in terms of the error used in the optimization process and in terms of COV). For complex ML models, the mean ratio of V_{true}/V_{pred} should be very close to 1.0 with a COV of 0.12, or less, and training and testing errors should be within a margin of ±20% of the converging error. Similar findings are expected for other similar applications with similar size databases.

Finally, none of the assessed existing code-oriented models meet the target performance requirements for a simplified shear strength model, which suggests there is room for improvement in code equation predictive performance. Also, none of the ML models analyzed in the literature review meet the performance requirements for complex ML models, which reflects the impact of improper training.

AUTHOR BIOS

Matías Rojas-León *is a Practicing Engineer. He received his BS, professional engineer degree, and MS from the University of Chile, Santiago, Chile; and his PhD in structural/earthquake engineering from the University of California, Los Angeles (UCLA), Los Angeles, CA. His research interests include seismic design and behavior of reinforced concrete structures, machine learning (ML) and statistical approaches in structural engineering, laboratory testing, and structures equipped with anti-seismic devices.*

Saman A. Abdullah *is a Research Scholar in the Department of Civil and Environmental Engineering at UCLA and a Lecturer in the Department of Civil Engineering, College of Engineering, at the University of Sulaimani, Sulaymaniyah, Kurdistan, Iraq. He is a member of ACI Committee 374, Performance-Based Seismic Design of Concrete Buildings, and ACI Subcommittees ACI 318-H, Seismic Provisions; 318-1W, Wind Provisions; and 369-F, Retrofit. His research interests include seismic and wind design of concrete structures and laboratory testing.*

Kristijan Kolozvari *is an Associate Professor in the Department of Civil and Environmental Engineering at California State University, Fullerton, Fullerton, CA. He is a member of ACI Committee 374, Performance-Based Seismic Design of Concrete Buildings. His research interests include developing and applying innovative analytical tools for nonlinear analysis of reinforced concrete structures, performance-based seismic design, seismic retrofit, tall building behavior and design, and earthquake resiliency.*

John W. Wallace, *FACI, is a Professor of civil engineering at UCLA. He is Chair of ACI Subcommittee 318-H, Seismic Provisions, and a member of ACI Committees 318, Structural Concrete Building Code; 369, Seismic Repair and Rehabilitation; and 374, Performance-Based Seismic Design of Concrete Buildings. His research interests include the response and design of buildings and bridges to earthquake actions, laboratory and field testing of structural components and systems, and seismic structural health monitoring.*

ACKNOWLEDGMENTS

M. Rojas-León acknowledges the support from the National Agency for Research and Development (ANID) Scholarship Program/DOCTORADO BECAS CHILE/2019-72200499. Any opinions, findings, and conclusions

expressed in this material are those of the authors and do not necessarily reflect those of the sponsors.

REFERENCES

Abdullah, S. A., 2019, "Reinforced Concrete Structural Walls: Test Database and Modeling Parameters," PhD dissertation, University of California, Los Angeles, Los Angeles, CA, 304 pp.

Abdullah, S. A., and Wallace, J. W., 2018, "UCLA-RC Walls Database for Reinforced Concrete Structural Walls," *Proceedings of the 11th National Conference in Earthquake Engineering*, Los Angeles, CA.

Abdullah, S. A., and Wallace, J. W., 2021, "New Nonlinear Modeling Parameters and Acceptance Criteria for RC Structural Walls," The 2021 Annual Conference of Los Angeles Tall Buildings Structural Design Council, Nov. 12, Los Angeles, CA.

ACI Committee 318, 2008, "Building Code Requirements for Structural Concrete (ACI 318-08) and Commentary (ACI 318R-08)," American Concrete Institute, Farmington Hills, MI, 473 pp.

ACI Committee 318, 2011, "Building Code Requirements for Structural Concrete (ACI 318-11) and Commentary (ACI 318R-11)," American Concrete Institute, Farmington Hills, MI, 503 pp.

ACI Committee 318, 2014, "Building Code Requirements for Structural Concrete (ACI 318-14) and Commentary (ACI 318R-14)," American Concrete Institute, Farmington Hills, MI, 520 pp.

ACI Committee 318, 2019, "Building Code Requirements for Structural Concrete (ACI 318-19) and Commentary (ACI 318R-19) (Reapproved 2022)," American Concrete Institute, Farmington Hills, MI, 624 pp.

AIJ, 1999, "Structural Design Guidelines for Reinforced Concrete Buildings," Architectural Institute of Japan, Tokyo, Japan.

Alzubi, J.; Nayyar, A.; and Kumar, A., 2018, "Machine Learning from Theory to Algorithms: An Overview," *Journal of Physics: Conference Series*, 2nd National Conference on Computational Intelligence (NCCI 2018), Bangalore, India, IOP Science, V. 1142, pp. 012012

Amazon Web Services, Inc., 2016, "Amazon Machine Learning Developer Guide," AWS, Seattle, WA.

ASCE/SEI 43-05, 2005, "Seismic Design Criteria for Structures, Systems, and Components in Nuclear Facilities," American Society of Civil Engineers, Reston, VA.

Barda, F.; Hanson, J. M.; and Corley, W. G., 1977, "Shear Strength of Low-Rise Walls with Boundary Elements," *Reinforced Concrete Structures in Seismic Zones*, SP-53, American Concrete Institute, Farmington Hills, MI, pp. 149-202.

Barret, J., 1974, "The Coefficient of Determination – Some Limitations," *The American Statistician*, V. 28, No. 1, pp. 19-20.

Breiman, L., 2001, "Random Forests," *Machine Learning*, V. 45, No. 1, pp. 5-32. doi: 10.1023/A:1010933404324

Bzdok, D.; Altman, N.; and Krzywinski, M., 2018, "Points of Significance: Statistics Versus Machine Learning," *Nature Methods*, V. 15, No. 4, pp. 233-234. doi: 10.1038/nmeth.4642

CSA A23.3-14, 2014, "Design of Concrete Structures," CSA Group, Toronto, ON, Canada.

Cardenas, A. E., and Magura, D. D., 1972, "Strength of High-Rise Shear Walls-Rectangular Cross Section," *Special Publication (American Philosophical Society)*, V. 36, pp. 119-150.

Carrillo, J., and Alcocer, S., 2013, "Shear Strength of Reinforced Concrete Walls for Seismic Design of Low-Rise Housing," *ACI Structural Journal*, V. 110, No. 3, May-June, pp. 415-426.

Chen, X. L.; Fu, J. P.; Yao, J. L.; and Gan, J. F., 2018, "Prediction of Shear Strength for Squat RC Walls Using a Hybrid ANN-PSO Model," *Engineering with Computers*, V. 34, No. 2, pp. 367-383. doi: 10.1007/s00366-017-0547-5

Dey, A., 2016, "Machine Learning Algorithms: A Review," *International Journal of Science and Research*, V. 7, No. 3, pp. 1174-1179.

EN 1998-1:2004, 2004, Eurocode 8, "Design of Structures for Earthquake Resistance, Part 1: General Rules, Seismic Actions and Rules for Buildings," European Committee for Standardization, Brussels, Belgium.

Feng, D. C.; Wang, W. J.; Mangalathu, S.; and Taciroglu, E., 2021, "Interpretable XGBoost-SHAP Machine-Learning Model for Shear Strength Prediction of Squat RC Walls," *Journal of Structural Engineering*, ASCE, V. 147, No. 11, p. 04021173. doi: 10.1061/(ASCE)ST.1943-541X.0003115

Gonfalonieri, A., 2019, "5 Ways to Deal with the Lack of Data in Machine Learning," KDnuggets, https://www.kdnuggets.com/2019/06/5-ways-lack-data-machine-learning.html. (last accessed Nov. 2, 2023)

Google Developers, 2022, "The Size and Quality of a Data Set," Google, Mountain View, CA, https://developers.google.com/machine-learning/data-prep/construct/collect/data-size-quality. (last accessed Nov. 2, 2023)

Gulec, C., and Whittaker, A. S., 2011, "Empirical Equations for Peak Shear Strength of Low Aspect Ratio Reinforced Concrete Walls," *ACI Structural Journal*, V. 108, No. 1, Jan.-Feb., pp. 80-89.

Gulec, C. K.; Whittaker, A. S.; and Bozidar Stojadinovic, B., 2009, "Peak Shear Strength of Squat Reinforced Concrete Walls with Boundary Barbells or Flanges," *ACI Structural Journal*, V. 106, No. 3, May-June, pp. 368-377.

Hoerl, A. E., and Kennard, R. W., 1970, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, V. 12, No. 1, pp. 55-67. doi: 10.1080/00401706.1970.10488634

Höge, M.; Wöhling, T.; and Nowak, W., 2018, "A Primer for Model Selection: The Decisive Role of Model Complexity," *Water Resources Research*, V. 54, No. 3, pp. 1688-1715. doi: 10.1002/2017WR021902

Hwang, S.-J., and Lee, H.-J., 2002, "Strength Prediction for Discontinuity Regions by Softened Strut-and-Tie Model," *Journal of Structural Engineering*, ASCE, V. 128, No. 12, pp. 1519-1526. doi: 10.1061/(ASCE)0733-9445(2002)128:12(1519)

Kassem, W., 2015, "Shear Strength of Squat Walls: A Strut-and-Tie Model and Closed-Form Design Formula," *Engineering Structures*, V. 84, pp. 430-438. doi: 10.1016/j.engstruct.2014.11.027

Keshtegar, B.; Nehdi, M. L.; Trung, N.-T.; and Kolahchi, R., 2021, "Predicting Load Capacity of Shear Walls Using SVR–RSM Model," *Applied Soft Computing*, V. 112, Nov., p. 107739. doi: 10.1016/j.asoc.2021.107739

Kim, J.-H., and Park, H.-G., 2020, "Shear and Shear-Friction Strengths of Squat Walls with Flanges," *ACI Structural Journal*, V. 117, No. 6, Nov., pp. 269-280. doi: 10.14359/51728075

Li, H., and Li, B., 2002, "Experimental Study on Seismic Restoring Performance of Reinforced Concrete Shear Walls," *Journal of Building Structures*, V. 32, No. 5, pp. 728-732.

Looi, D. T. W., and Su, R. K. L., 2017, "Predictive Seismic Shear Capacity of Rectangular Squat RC Shear Walls in Flexural and Shear Zones," 16th World Conference on Earthquake Engineering, Santiago, Chile, pp. 9-13.

Moradi, M., and Hariri-Ardebili, M., 2019, "Developing a Library of Shear Walls Database and the Neural Network Based Predictive Meta-Model," *Applied Sciences*, V. 9, No. 12, p. 2562. doi: 10.3390/app9122562

Morgan, N., and Bourlard, H., 1989, "Generalization and Parameter Estimation in Feedforward Nets: Some Experiments," *Advances in Neural Information Processing Systems*, pp. 630-637

NZS 3101-06, 1995, "New Zealand Concrete Structures Standards," Standards New Zealand, Wellington, New Zealand.

Rojas-León, M., 2022, "Framework to Set Model Performance Requirements Applied to the RC Wall Shear Strength Problem and Proposition of New Code-Oriented Equation," PhD dissertation, University of California, Los Angeles, Los Angeles, CA, 260 pp.

Sánchez-Alejandre, A., and Alcocer, S., 2010, "Shear Strength of Squat Reinforced Concrete Walls Subjected to Earthquake Loading – Trends and Models," *Engineering Structures*, V. 32, No. 8, pp. 2466-2476. doi: 10.1016/j.engstruct.2010.04.022

Segal, M., and Xiao, Y., 2011, "Multivariate Random Forests," *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery*, V. 1, No. 1, pp. 80-87. doi: 10.1002/widm.12

Segura, C. L., and Wallace, J. W., 2018, "Seismic Performance Limitations and Detailing of Slender Reinforced Concrete Walls," *ACI Structural Journal*, V. 115, No. 3, May, pp. 849-859. doi: 10.14359/51701918

Tanaka, J. S., 1987, "How Big Is Big Enough?: Sample Size and Goodness of Fit in Structural Equation Models with Latent Variables," *Child Development*, V. 58, No. 1, pp. 134-146. doi: 10.2307/1130296

Tibshirani, R., 1996, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, V. 58, No. 1, pp. 267-288.

Wood, S. L., 1990, "Shear Strength of Low-Rise Reinforced Concrete Walls," *ACI Structural Journal*, V. 87, No. 1, Jan.-Feb., pp. 99-107.

Zhang, C., and Ma, Y., 2012, *Ensemble Machine Learning: Methods and Applications*, Springer.

Zheng, S.; Qin, Q.; Yang, W.; Gan, C.; Zhang, Y.; and Ding, S., 2015, "Experimental Research on the Seismic Behaviors of Squat RC Shear Walls Under Offshore Atmospheric Environment," *Journal of Harbin Institute of Technology*, V. 47, No. 12, pp. 64-69. (in Chinese)

Zou, H., and Hastie, T., 2005, "Regularization and Variable Selection via Elastic Net," *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, V. 67, No. 2, pp. 301-320. doi: 10.1111/j.1467-9868.2005.00503.x

APPENDIX

Table A.1 – Existing wall shear strength models in building codes and standards

Model	Comments
ACI 318-19, Section 18.10 $V_n = A_{cv}\left(\alpha_c \lambda \sqrt{f_c'} + \rho_{wh} f_{ywh}\right) \le 0.83 A_{cv}\sqrt{f_c'}$ $\alpha_c = \begin{cases} 0.25 & \text{for } h_w/l_w \le 1.5 \\ 0.17 & \text{for } h_w/l_w \ge 2.0 \\ 0.25 - 0.17\left(\frac{h_w}{l_w} - 1.5\right) & \text{for } 1.5 < h_w/l_w < 2.0 \end{cases}$ λ = 1.0 for normal-weight concrete. For lightweight concrete, it ranges from 0.75 to 0.85 depending on the composition of the aggregates.	Wall shear strength is determined using Eq. (18.10.4.1). The upper limit of $0.83A_{cv}\sqrt{f_c'}$ (for an individual wall) is intended to prevent diagonal compression failure and has been in the ACI 318 Code since 1971. For walls with $h_w/l_w \le 2.0$ it is required that ρ_v is no less than ρ_h.
EC8-2004 $V_n = v_n t_w d$ $v_n = \begin{cases} \rho_h f_{yh}\left(\frac{M_u}{V_u l_w} - 0.3\right) + \rho_v f_{yv}\left(1.3 - \frac{M_u}{V_u l_w}\right) & \text{for } \frac{1.5P_u}{A_t f_c'} < 0.1 \\ 0.15\sqrt{f_c'} + \left[\rho_h f_{yh}\left(\frac{M_u}{V_u l_w} - 0.3\right) + \rho_v f_{yv}\left(1.3 - \frac{M_u}{V_u l_w}\right)\right] & \text{for } \frac{1.5P_u}{A_t f_c'} \ge 0.1 \end{cases}$	Concrete contribution is ignored for walls subjected to low axial stresses ($1.5P_u/(A_t f_c') < 0.1$). Wall shear strength depends on vertical and horizontal web reinforcement and applied moment-to-shear ratio.
NZS 3101-2006 $V_n = V_c + V_s$ $V_s = \frac{A_v f_{ywh} d}{s}$, $\frac{V_c}{A_{cv}} = \begin{cases} min\left\{0.17\sqrt{f_c'}, 0.17\left(\sqrt{f_c'} + \frac{P_u}{A_g}\right)\right\} & \text{Simplified Method} \\ min\left\{\left(0.27\sqrt{f_c'} + \frac{P_u}{4A_g}\right), \left(0.05\sqrt{f_c'} + \frac{l_w\left(0.1\sqrt{f_c'} + 0.2\frac{P_u}{A_g}\right)}{\frac{M_u}{V_u} - \frac{l_w}{2}}\right)\right\} & \text{Detailed Method} \end{cases}$	The simplified method may only be used when the vertical reinforcement ratio along the entire wall exceeds 0.003, and the reinforcement spacing does not exceed 300 mm (12 in) in any direction. The detailed equation does not apply if $\left(\frac{M_u}{V_u} - \frac{l_w}{2}\right) \le 0$.
AIJ-1999 $V_n = V_c + V_s$ $V_c = \frac{tan(\theta)(1-\beta) t_w l_w \nu f_c'}{2} \ge 0$, $V_s = \rho_{wh} f_{ywh} t_w l_w cot(\xi)$, $tan(\theta) = \sqrt{\left(\frac{h_w}{l_w}\right)^2 + 1} - \frac{h_w}{l_w}$ $\nu = 0.7 - \frac{f_c'}{2000}$, $\beta = \frac{\left(1 + cot^2(\xi)\right)\rho_{wh} f_{ywh}}{\nu f_c'}$, $cot(\xi) = 1$ (for truss mechanisms)	In this model, besides the truss analogy, shear is assumed to be resisted through an arch mechanism. The contribution of the arch mechanism decreases with the amount of the web horizontal steel.
ASCE/SEI 43-05 $V_n = \left[0.70\sqrt{f_c'} - 0.28\sqrt{f_c'}\left(\frac{h_w}{l_w} - 0.5\right) + \frac{P_u}{4 l_w t_w} + \rho_{se} f_{y1} \le 20\sqrt{f_c'}\right] t_w d_w \le 1.66 A_{cv}\sqrt{f_c'}$ $d_w = 0.6 l_w$; $\rho_{se} = A\rho_v + B\rho_h$ $A = \begin{cases} 1 & \text{for } \frac{h_w}{l_w} \le 0.5 \\ 1.5 - \frac{h_w}{l_w} & \text{for } 0.5 < \frac{h_w}{l_w} \le 1.5 \\ 0 & \text{for } \frac{h_w}{l_w} > 1.5 \end{cases}$, $B = \begin{cases} 1 & \text{for } \frac{h_w}{l_w} \le 0.5 \\ \frac{h_w}{l_w} - 0.5 & \text{for } 0.5 < \frac{h_w}{l_w} \le 1.5 \\ 0 & \text{for } \frac{h_w}{l_w} > 1.5 \end{cases}$	ASCE/SEI 43-05 adopted the work done by Barda et al. (1977), with modifications to extend its applicability. This equation is meant to predict the peak shear strength of walls with barbells or flanges, common in nuclear power plants, and applies to walls with $\frac{h_w}{l_w} \le 2.0$ and vertical and horizontal web reinforcement ratios ≤ 1%. If the reinforcement ratios exceed 1%, the combined reinforcement ratio ρ_{se} is limited to 1%.

*All equations are in units of MPa, mm, and kN

Table A.2 – Existing wall shear strength models reported in the literature

Model	Database	Comments
Barda et al. (1977) $V_n = \left[0.66\sqrt{f_c'} - 0.21\sqrt{f_c'}\frac{h_w}{l_w} + \frac{P_u}{4l_w t_w} + \rho_v f_{yv}\right] t_w d$	• 8 flanged walls • $\frac{h_w}{l_w}$ between 0.25 – 1.00 • ρ_{be} between 1.8% - 6.4% • ρ_{wh} and ρ_{wv} between 0.0% - 0.5%	It is meant to predict the peak shear strength of walls in low-rise buildings.
Wood (1990) $0.5A_{cv}\sqrt{f_c'} \le V_n = \frac{A_{vf}f_y}{4} \le 0.83A_{cv}\sqrt{f_c'}$ $A_{vf}f_y = 2A_{be}f_{ybe} + A_{wv}f_{ywv}$	• 143 squat walls reported to have failed in shear • ~105 barbell, ~20 flanged walls, and ~18 rectangular cross-sections. • $\frac{M_u}{V_u l_w} \le 2.0$, with $0.5 \le \frac{M_u}{V_u l_w} \le 1.0$ for more than 75% of the test specimens. • $0.7 \le \frac{P_u}{A_g f_c'} \le 0.18$ for 18 specimens, and $\frac{P_u}{A_g f_c'} = 0$ for the rest.	The model does not consider the concrete and steel contributions as two different terms. Instead, it only uses the concrete contribution to define a lower and upper limit.
Sánchez-Alejandre and Alcocer (2010) $V_n = \left(\left(\gamma\eta_v + 0.04\frac{P_u}{A_g}\right)\sqrt{f_c'} + \eta_h\rho_h f_{yh}\right)A_g$ $\gamma = 0.42 - 0.08\frac{M_u}{V_u l_w}$ or $\gamma = 0.42 - 0.08(\%R_{max})$ $\eta_v = 0.75 + 0.05\rho_v f_{yv}$ $\eta_h = 1 - 0.16\rho_h f_{yh} \ge 0.20$ R_{max} = drift angle	• 80 rectangular walls with diagonal tension failure mode • Most of the walls have $\frac{M_u}{V_u l_w} \le 1.0$ • Low web reinforcement ratios and axial loads. • Drift angles < 1%.	The model depends on the amount of web reinforcement (hor. and ver.) that has reached plastic strains at a given drift angle (R_{max}), and although the γ factor depends on R_{max}, it can also be expressed in terms of $M_u/(V_u l_w)$.
Gulec and Whittaker (2011) $V_n = \begin{cases} \text{For rectangular walls:} \\ V_{rec} \le 0.83\sqrt{f_c'}A_{cv} \\ \text{For barbell/flanged walls with } \frac{A_t}{A_{cv}} \ge 1.25\text{:} \\ V_{be} \le 1.25\sqrt{f_c'}A_g \\ \text{For barbell/flanged walls with } \frac{A_t}{A_{cv}} < 1.25\text{:} \\ min(V_{rec}, V_{be}) \le 10\sqrt{f_c'}A_{cv} \end{cases}$ $V_{rec} = \frac{0.83\sqrt{f_c'}A_{cv} + 0.25F_{vw} + 0.20F_{vbe} + 0.40P_u}{\sqrt{\frac{h_w}{l_w}}}$ $V_{be} = \frac{0.04f_c'A_{\text{eff}} + 0.40F_{vw} + 0.15F_{vbe} + 0.35P_u}{\sqrt{\frac{h_w}{l_w}}}$	• Cantilever walls • One database of 74 rectangular walls • Second database of 153 walls (79 barbell walls and 74 flanged walls) • $\frac{h_w}{l_w}$ between 0.25 – 2.0 • f_c' between 13.7 – 51.0 MPa (2,000 – 7,400 psi) • $\frac{P_u}{A_t f_c'}$ between 0 and 0.14 • $\rho_{wh}f_{ywh}$ between 0 – 5.8 MPa (835 psi) • $\rho_{wv}f_{ywv}$ between 0 – 12.8 MPa (1,860 psi) • $\rho_{be*}f_{ybe}$ between 0 – 14.1 MPa (2,050 psi)	The model is based on a free-body diagram of a low aspect ratio wall with an inclined (shear) crack. The boundary element reinforcement (ρ_{be*}) is calculated as $2A_{s,be}/A_t$, where $A_{s,be}$ is the area of vertical reinforcement in each BE. Specific parameters for this model are: F_{vw} = force attributed to vertical web reinforcement, and F_{vbe} = force attributed to both BEs reinforcement.
Carrillo and Alcocer (2013) $V_n = \left(\alpha_1\sqrt{f_c'} + \eta_h\rho_h f_{yh}\right)A_{cv} \le \alpha_2\sqrt{f_c'}A_{cv}$ $\eta_h = \begin{cases} 0.8 \ for\ deformed\ bars \\ 0.7 \ for\ welded - wire\ mesh \end{cases}$ $\alpha_1 = 0.21 - 0.02\left(\frac{M_u}{V_u l_w}\right)$ $\alpha_2 = 0.40$	• 39 walls from quasi-static and shake-table tests. • $\frac{h_w}{l_w} \approx 0.5, 1.0, 2.0$ • Normal-weight, lightweight, and self-consolidating concrete. • ρ_{wh} and ρ_{wv} between 0.00% - 0.28% • ρ_{be} between 0.22% - 1.50%	Meant to be used for walls in typical low-rise housing in Latin America (low concrete strength and wall thickness of ~4 in.)
Kassem (2015) $V_n = v_n t_w d_w$ $v_n =$ $\begin{cases} \text{For rectangular walls:} \\ 0.44f_c'\left[\psi k_s sin(2\alpha) + 0.10\omega_h\frac{h_w}{d_w} + 0.30\omega_v cot(\alpha)\right] \\ \text{For flanged walls:} \\ 0.67f_c'\left[\psi k_s sin(2\alpha) + 0.16\omega_h\frac{h_w}{d_w} + 1.74\omega_v cot(\alpha)\right] \end{cases}$ $\psi = 0.95 - \frac{f_c'}{250}$; $k_s = \frac{a_s}{d_w}$; $\alpha = tan^{-1}\left(\frac{h_w}{d_w}\right)$ $d_w = d - \frac{a_s}{3}$; $\omega_h = \frac{\rho_h f_{yh}}{f_c'}$; $\omega_v = \frac{\rho_v f_{yv}}{f_c'}$ $a_s = \left(0.25 + 0.85\frac{P_u}{A_{cv}f_c'}\right)l_w$	• Cantilever walls 287 rectangular walls • $\frac{h_w}{l_w}$ between 0.25 – 1.0 • $\frac{P_u}{A_g f_c'}$ between 0.00 – 0.23 • ρ_{wh} between 0.00% - 1.61% • ρ_{wv} between 0.00% - 2.87% 358 flanged walls • $\frac{h_w}{l_w}$ between 0.21 – 1.60 • $\frac{P_u}{A_g f_c'}$ between 0.00 – 0.34 • ρ_{wh} between 0.00% - 2.89% • ρ_{wv} between 0.00% - 2.89%	It is a mathematical equation based on the strut-and-tie model. Since flanged walls are more susceptible to diagonal compression failure than rectangular walls, two separate databases of rectangular and flanged walls were used to calibrate the equations. The equations are using the coefficients before including variations to account for safety factors. The length of compression zone (a_s) can be determined from sectional analysis instead.
Looi and Su (2017) $V_n = v_n A_g f_c'$	• ~ 150 rectangular walls with shear and flexure-shear failure modes	The proposed model is entirely based on a multi-parameter

$\frac{v_n}{f_c'} = 0.034 + A\left(\frac{P_u}{A_g f_c'}\right)^{1.3} + B\rho_v \frac{f_{yv}}{f_c'} + C\rho_h \frac{f_{yh}}{f_c'} + D\rho_{be} \frac{f_{ybe}}{f_{cc}'} \leq 0.24$ f_{cc}': confined concrete strength $A = 0.283 - 0.084\frac{M_u}{V_u d}$; $B = 0.4 - 0.15\frac{M_u}{V_u d}$ $C = 0.5 - 0.2\frac{M_u}{V_u d}$; $D = -0.08 + 0.06\frac{M_u}{V_u d}$	• f_c' between 15.7 – 70.3 MPa • $\frac{P_u}{A_g f_c'}$ between 0.00 – 0.40 • $\frac{M_u}{V_u l_w}$ between 0.4 – 2.6 • ρ_{wh} between 0.11% - 1.72% • ρ_{wv} between 0.13% - 2.84% • ρ_{be} between 0.00% - 13.46%	regression, with relevant parameters based on a literature review. The coefficients in the equation do not have units; therefore, they are the same no matter the set of units used as long as they are consistent.

Table A.3 – Predictive RC wall shear strength models obtained with machine learning

Chen et al. (2018)	Moradi and Hariri-Ardebili (2019)	Keshtegar et al. (2021)
ANN-PSO (Artificial Neural Network implemented with Particle Swarm Optimization algorithm) with 1 hidden layer and 13 neurons	**ANN** with 13 input parameters, 1 hidden layer, and 10 neurons	**SVR-RSM** (Support Vector Regression coupled with Response Surface Model)
• 6 input variables h_w/l_w, P_u, f_c', A_{cv}, $\rho_h f_y$, $\rho_v f_y$	• 13 input variables P_u , h_w , l_w , t_w , b_f , t_f , ρ_{wh} , ρ_{wv} , f_y , f_c' , "vertical column reinforcement ratio", and "horizontal column reinforcement ratio".	• 15 input variables: f_c', ρ_{wh}, ρ_{vw}, f_{ywh}, f_{ywv}, b_f, t_f, t_w, h_w, h_w/l_w , P_u , A_{tot} , "effective length of wall," "longitudinal reinforcement ratio of flanges", "yield strength of bars in flanges".
• Output variable V_{true}	• Output variables V_{true}, lateral in-plane stiffness, drift ratio	• Output variable: V_{true}
• Database – 139 tests – 80% for training test and 20% for testing set – Rectangular walls – h_w/l_w between 0.25 – 2.0 – $\frac{P_u}{A_g f_c'}$ between 0.0 – 0.35 – ρ_h between 0.0 – 1.96 % – ρ_v between 0.0 – 2.93 %	• Database – 329 tests – 85% for training and 15% for testing set – ρ_{wh} between 0 – 6.69% – ρ_{wv} between 0 – 14.33% – "Horizontal column reinforcement ratio" between 0 – 6.69% – "Vertical column reinforcement ratio" between 0 – 14.33% – V_{true} between 15.42 – 3,231kN	• Database – 208 tests – 70% for training test and 30% for testing set – $\frac{h_w}{l_w}$ between 0.21 – 2.4 – ρ_{wh} between 0.00% - 2.44% – ρ_{wv} between 0.00% - 2.90% – V_{true} between 70 – 2,483kN
• Error indicator used in the training – Root mean square error (RMSE) – Coefficient of determination (R^2)	• Error indicator used in the training – Mean square error (MSE)	• Error indicator used in the training – Root mean square error (RMSE)