11^{th} International Conference on Fracture Mechanics of Concrete and Concrete Structures
FraMCoS-11
J. M. Chandra Kishen, A. Ramaswamy, S. Ray and R. Vidyasagar (Eds)
https://doi.org/10.21012/FC11.092353

MACHINE LEARNING REGRESSION MODELS FOR PEAK SHEAR STRENGTH PREDICTION OF SQUAT SHEAR WALLS

SHASHANK TYAGI*, **APPA RAO G.**†

*Indian Institute of Technology Madras
Chennai, India
e-mail: shashank.tyagi444@gmail.com

†Indian Institute of Technology Madras
Chennai, India
e-mail: garao@iitm.ac.in

Key words: Reinforced Concrete, Squat Structural Walls, Machine Learning,Regression Analysis,Earthquake Engineering

Abstract. A Reinforced Concrete wall is designed to contribute to most if not the whole lateral load carrying capacity of the structure. The lateral load carrying capacity of such wall is indicated by the peak shear strength of the walls. The strength of Slender Walls can be found fairly accurately by section analysis which has been studied quite extensively and has been reported by most of the codes.However the codes fail to provide an accurate method for determining the peak shear strength of Squat Walls. Various theories based on mechanics developed to study the behavior of concrete have been applied to structural walls which give accurate results but are time-consuming and resource-heavy. This complexity in peak shear strength calculation has been attributed to various parameters affecting the behavior of such walls dominated by shear. Machine learning models Polynomial Regression, KNN Regression, Decision Tree Regression, Random Forest and Boosting Method have been done on a database of 594 Squat Structural Walls.The accuracy of these models has been reported and the importance of parameters has been found. The model is flexible to add more data from the results obtained from the experimental studies to be carried out in the future

1 INTRODUCTION

Structural Walls systems are designed to carry the lateral load on the building and expected to follow earthquake design philosophy [1]. In recent earthquakes, we have seen failures due to construction errors and a lack of detailing which ensure ductile failure in the walls. These walls are to be designed for sufficient strength to effectively take the load coming on the structure during earthquakes.

When a wall experiences flexural-dominated failure, it exhibits cantilever action and causes the vertical reinforcement at the ends to yield. The strength, in this case, can be obtained by performing section analysis which has been verified by past experiments and hence has been incorporated in building codes worldwide. Determining the peak strength of Squat Walls is difficult due to the shear-dominated failure mode, which causes brittle failure and a sudden loss of strength beyond a specific displacement. Many experimental studies [2] has been carried over the years and various codes and researchers [3] have provided equation for peak strength determination. However, discrepancies exist among the suggested equations [4], as the parameters considered in the studies are not consis-

tent. Various finite element studies have also been conducted for the same and have given good estimates of the wall capacity [5]. While studying wall behavior through this method is crucial and provides valuable insights, its time-consuming and resource-intensive nature makes it challenging to analyze a portfolio of structures. The continuing improvement in machine learning can be a viable option to better understand the contribution of various parameters and prediction of shear strength in squat structural walls.
Mangalathu et al. (2020) employed data-driven machine learning models to identify failure modes of shear walls. Similarly,Kiani et al. (2019) applied machine learning methods to develop seismic fragility curves for predicting potential damages following earthquakes. Mangalathu and Jeon utilized machine learning techniques for failure mode classification and shear strength prediction in reinforced concrete beam-column joints. Additionally, Ahmed Faleh Al-Bayati utilized Artificial Neural Networks (ANN) for predicting the shear strength of Squat Shear Walls (Al-Bayati, 2023).

2 Experimental Database

The performance of a machine learning model depends on the diversification in the database under study. This enables the regression models to identify the key parameters . A detailed description of the collected database and insights into the major parameters is discussed in this section.

2.1 Collection of Squat Reinforced Wall Database

The current study considers a database comprising experimental studies of 614 structural walls comprising of different cross sections. The database in the current study was extracted using the database provided by Chetchotisak, P. et al. [6] and the NEES Database repository. On the basis of a cross-section of the wall either categorized as rectangular, rectangular with boundary elements, barbell, or flanged section wall. Filtration is applied to the database to satisfy the following criterion-

1. All walls should have an aspect ratio less than 2.0
2. The minimum length of the wall should be 500mm.
3. The minimum height of the wall should be 500mm.
4. The thickness of the walls should be at least 60mm.

After the filtration, the database was reduced to 591 walls comprising 73 rectangular, 189 rectangular sections with boundary elements, 205 barbell, and 124 flanged section walls.Table 1 shows the mean and standard deviation of major parameters in the dataset after filtration.

Table 1: Mean and Standard Deviation of Major Parameters

Parameter	Mean	Standard Deviation
Aspect Ratio	0.92	0.42
l_w/t_w	17.73	9.77
f_{ck}	35.08	18.19
ρ_h	0.65	0.45
ρ_v	0.69	0.50
ρ_b	3.01	2.24
f_{yh}	443.76	145.53
f_{yv}	452.60	136.19
f_{yb}	447.93	121.86
A_b/A_g	0.10	0.11
$N/f_{ck}A_g$	0.045	0.067

The histogram of major parameters in the dataset is shown in Figure 1.Aspect Ratio (AR) is defined as the ratio of height to the length of the wall. lw/tw represents the ratio of length to the thickness of the wall.f_{ck} represents the concrete compressive strength in MPa.Horizontal Reinforcement Ratio (ρ_h) represents the horizontal reinforcement ratio defined as the ratio of total horizontal reinforcement area in the wall to the area along the thickness of the web. Vertical

Reinforcement Ratio(ρ_v) is the ratio of total vertical reinforcement in the wall to the wall cross sectional area. Boundary Element ratio(ρ_b) is defined as the ratio of vertical reinforcement in the boundary to the total boundary element area of the wall.

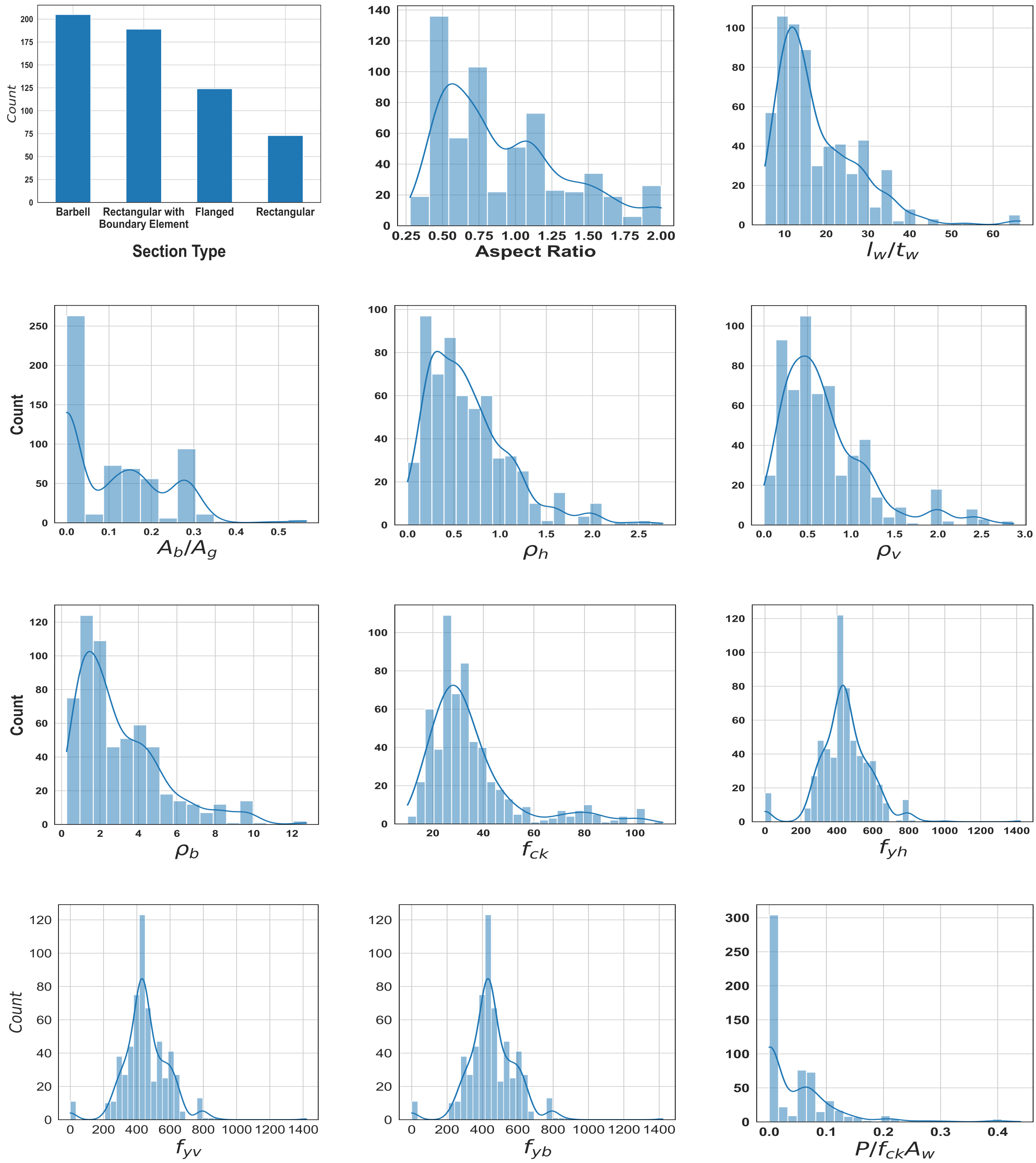


Figure 1: Histogram with KDE plot of major paramaters

f_{yh} and f_{yv} are the yield strength of horizontal reinforcement, vertical reinforcement in the wall in MPa.f_{yb} is the yield strength of vertical steel provided in boundary element of the wall in MPa.A_b/A_g is the ratio of the area of boundary element to the gross area of the wall. This ratio represents the contribution of the boundary element to the total area of the wall. Axial Load Ratio($N/f_{ck}A_w$) is defined as the ratio of Axial Load,N (in kN) applied to the product of concrete compressive strength (f_{ck}) and gross area of the wall(A_g)Figure 1 below shows the histogram of various wall parameters considered in the study.
No external moments were applied to the walls. It was found that no axial load was applied for 211 walls. The strength of horizontal and vertical wall reinforcement is found to be similar in most cases(358). From the literature, it has been found that the following are considered the major parameters influencing the shear strength of the squat structural wall.

- Material Properties - f_{ck},f_{yh},f_{yv},f_{yb}
- Geometric Wall Properties - l_w,h_w,t_w,AR
- Reinforcement Quantity - ρ_v,ρ_h,ρ_b
- Axial Load Ratio - $P/f_{ck}A_g$

3 Input Parameters

The identification of suitable input parameters is crucial for the execution of a machine-learning model. These parameters should encompass the major factors and establish a meaningful relationship among them. However, the presence of diverse units among the major parameters can introduce scale sensitivity and convergence issues. To address this, a set of input parameters has been defined based on a literature study by Mangalathu S. et al [7]. .For the current study,input parameters for the learning models have been selected which are shown in Table 2 .The last parameter P8 is a set of 4 columns created for categorical classification for the section type

Table 2: Input Parameters

Input Parameter Label	Parameter
P1	AR
P2	l_w/t_w
P3	A_b/A_g
P4	$f_{yh}\rho_h/f_{ck}$
P5	$f_{yv}\rho_v/f_{ck}$
P6	$f_{yb}\rho_b/f_{ck}$
P7	$N/f_{ck}A_g$
P8	Section Type

In the given context, the first four parameters pertain to the walls' geometric characteristics. On the other hand, P5 to P7 represents the reinforcement index, which quantifies the ratio of the multiplication between the provided reinforcement ratio and the yield strength of the corresponding reinforcement, normalized by the concrete strength of the wall. Parameter P8 considers the influence of axial load on the wall.

4 Brief Summary of Regression Techniques in Machine Learning

It is important to note that each regression method has its own unique approach, which can yield varying results depending on the distribution and relationship between the input parameters and the target variable. The choice of regression method should be carefully considered based on these factors to ensure optimal performance. The models used in the study are Polynomial Regression Model, kNN Regression Model, Decision Tree Regression Model, Random Forest Regression, Boosting Methods(ADABoost, CATBoost, LightGBM, XGBoost)In this section, we will provide an overview of these regression models-

4.1 Polynomial Regression Analysis

Polynomial Regression tries to fit a line of degree n, as the degree of the fitted curve increases the flexibility, which enables the model fits better to the training data, but this also makes it more susceptible to overfitting [8].

4.2 KNN Regression Analysis

KNN regression is a non parametric supervised distance based algorithm. It stores all the data points in the training dataset and calculates its distance from the observed point. Then it identifies n nearest point to the observed point and returns output as the average of values of the identified neighbors. As the values of the observed neighbors decrease kNN decision boundary becomes more flexible [8]. The Euclidean formulae for finding the distance to the nearest point is given below.

$$d(x, y) = \sum_{i=1}^{n} \sqrt{(x^i - y^i)^2} \tag{1}$$

where x^i and y^i are the difference in coordinates of the observed data point to the i^{th} point in the training dataset.

4.3 Decision Trees Regression Model

Decision tree regression is a supervised learning algorithm used for predicting continuous target variables [9]. It builds a tree-like model of decisions by considering the features of the input data. The splitting of the tree is determined to enhance homogeneity and maximize information gain. Starting from the root node, the tree is constructed recursively, dividing the data based on selected input parameters and their corresponding threshold values. The objective is to identify the most informative input feature at each split. Additionally, pruning techniques are applied to mitigate overfitting to the training dataset.

4.4 Random Forest Regression Model

Random Forest Regression is an ensemble technique created by Tim Kan Ho in 1995. It works by creating an array of decision trees during the training of data. It selects each split using a feature bagging technique that selects a random subset of features at each candidate split. that For the Regression tasks, the average prediction of the decision trees is returned, this reduces the variance due to the overfitting to the training data.

4.5 Boosting Methods

Boosting methods in machine learning aim to minimize prediction errors by sequentially training multiple models. This iterative approach improves the overall accuracy of the system by addressing the shortcomings of individual models [10]. The specific boosting methods employed in this study will now be described.

4.5.1 ADABoost

ADABoost is an ensemble algorithm that trains weak learners on weighted data [11]. It starts by predicting the original dataset with equal weights. In each iteration, it assigns higher weights to points with higher prediction errors. AdaBoost continues adding learners until reaching a limit on models or accuracy.

4.5.2 XGBoost

Extreme Gradient Boosting (XGBoost) is an exceptionally effective boosting algorithm designed for ensemble tree models. It employs regularization techniques to address complex tree structures [12]. However, it is worth noting that XGBoost can be computationally and memory intensive due to its boosting methodology.

4.5.3 LightGBM

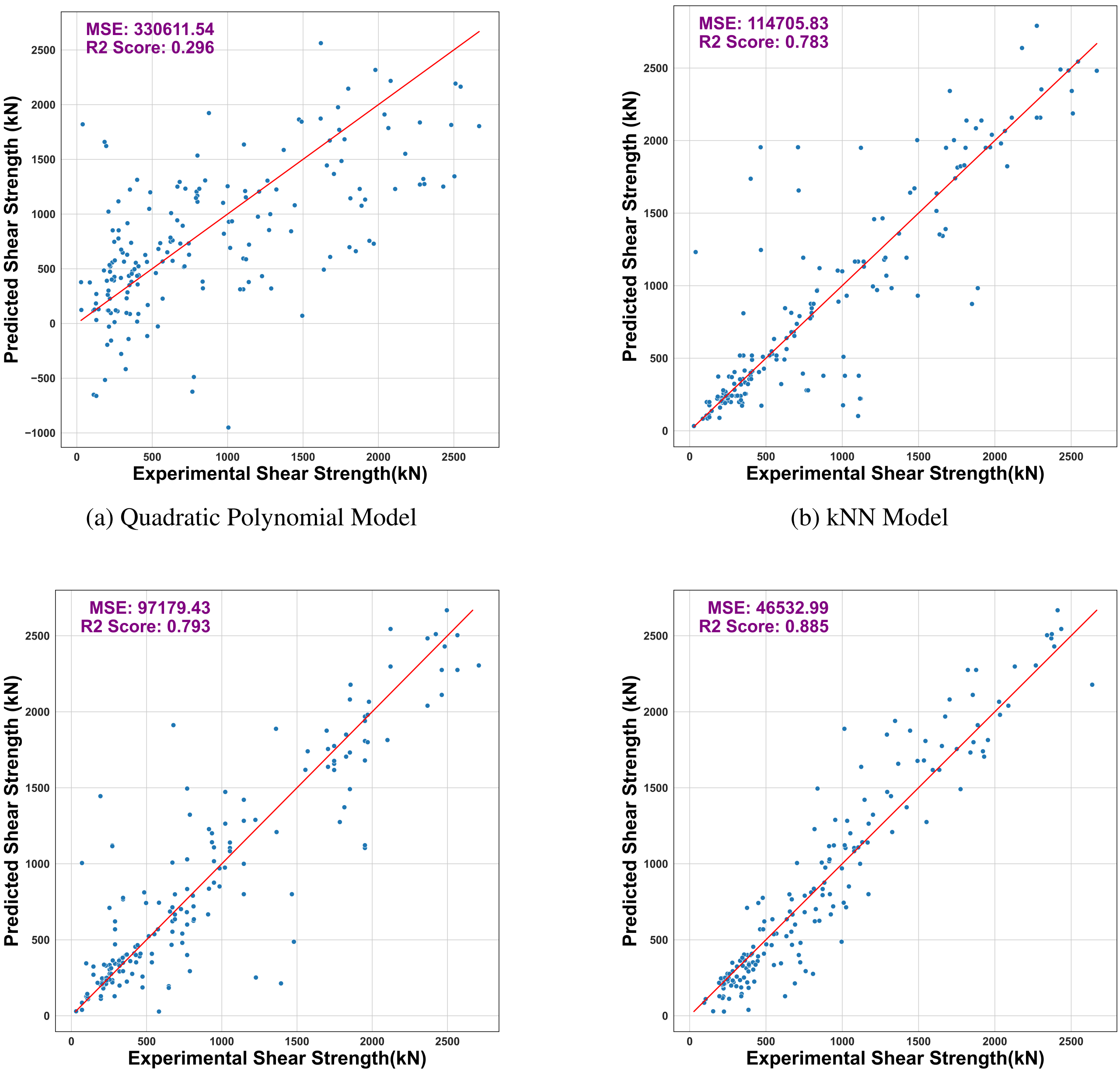
Light GBM is a high-speed boosting algorithm that utilizes a decision tree model. Unlike traditional depth-wise growth, it grows the tree in a leaf-wise manner. It selects the leaf with the maximum delta loss to expand further. By employing this leaf-wise approach, Light GBM can achieve greater loss reduction compared to level-wise algorithms when growing the same leaf.

4.5.4 CATBoost

Categorical Boosting(CAT Boost) is an ensemble tree based algorithm developed to handle categorical features without manual prepossessing [13]. It has automatic scaling features and uses regularization techniques to prevent overfitting.

5 Prediction to Shear Strength through Machine Learning Models

The machine learning techniques described in the previous section will be used to predict the shear strength of the wall.sd The characteristics of the wall that were described initially were converted to eight input parameters representing the geometry, reinforcement design, and axial load present on the wall. The models discussed in the previous section were based on Python open-source library sci-kit learn.



(a) Quadratic Polynomial Model

(b) kNN Model

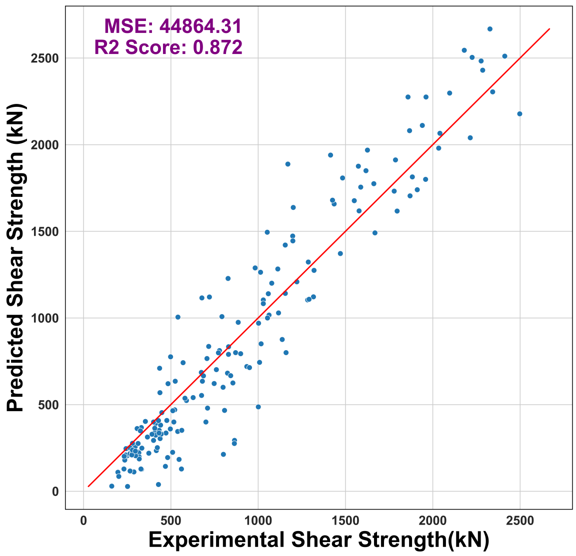
(c) Decision Tree Model

(d) Random Forest Model

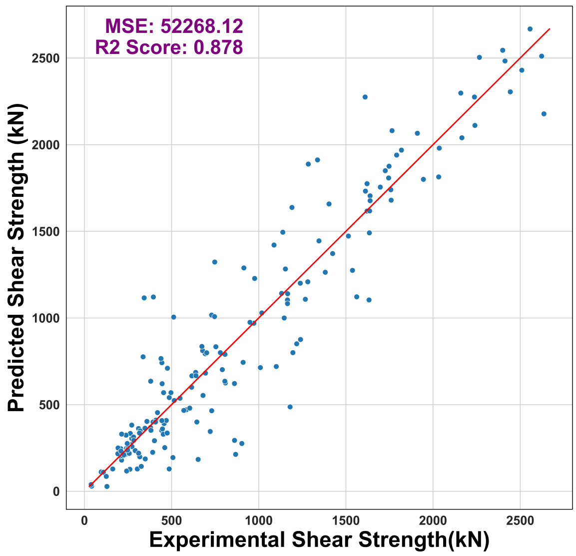
Figure 2: Experimental vs Predict Shear Strength of first four models

.The complete dataset is shuffled and divided into two sets of training and test data. The training data consist of 70% of the total data and will be used to train various ML models. Then the model performance is evaluated based on its accuracy in predicting strength of the test data independently. The performance of the model will be evaluated using mean squared error and R2 score. Most of the model cost functions are implemented to minimize the mean squared error. The performance of ML models is shown in Figure 3

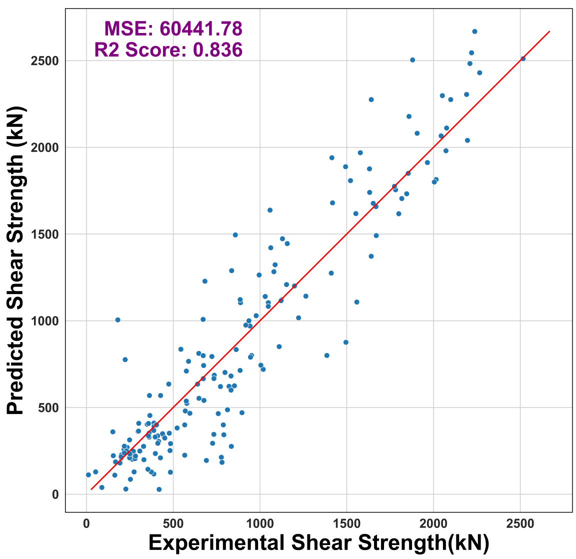
Initially, a Linear Regression analysis was carried out which gave a very high MSE and low bias, so a 2^{nd}-degree polynomial was chosen which improved the performance of the model and is shown in Figure 2a. A 3^{rd} or higher degree polynomial was overfitted and had high variance resulting in huge error while predicting.



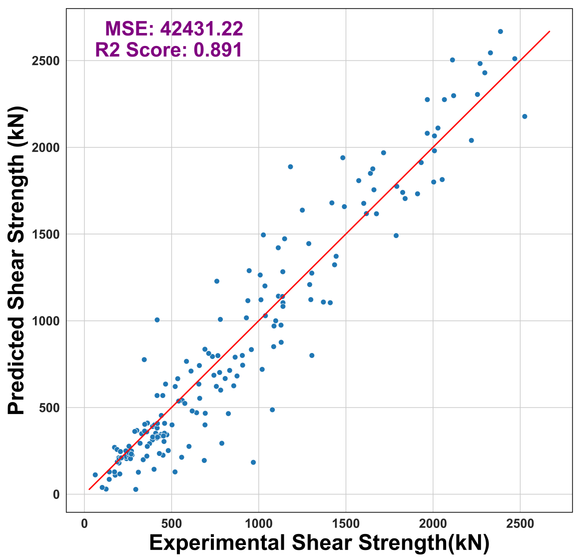
(a) ADA Boost



(b) XGBoost



(c) Light GBM Boost



(d) CAT Boost

Figure 3: Experimental vs Predict Shear Strength of Boosting Methods

As in Figure 2b, KNN showed improvement in prediction by reducing overall MSE and improving overall R2 score. It gave satisfactory results for walls with lower strength and showed high variability for walls with higher strength. The decision tree model did not exhibit significant improvement and demonstrated a high level of variability, which will be addressed through the use of boosting methods. By employing a random forest model, the accuracy of predictions improved substantially, effectively mitigating the variability observed in previous models. As previously discussed, boosting methods identify crucial parameters and employ regularization techniques to prevent overfitting to the training set. All boosting methods demonstrated improvement in terms of both MSE and R2 scores, effectively reducing variance. Using the random forest model, the importance of the input parameters was determined shown in Figure 4.

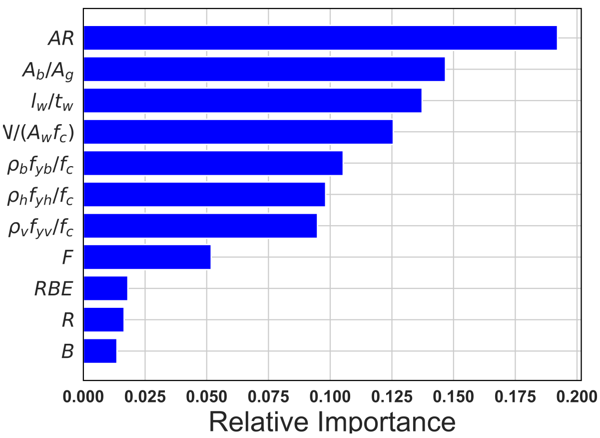


Figure 4: Importance of Input Parameters

It was found that the Aspect ratio is the most important parameter while making decisions for the tree split. It was followed by the A_b/A_g and then l_w/t_w. Out of the three reinforcement index for the design of the wall, the boundary element index was found to be the most important. Horizontal and Vertical reinforcement index was found to be equally influencing the peak shear strength. The Axial load ratio also increased the shear strength capacity of the walls.

6 Conclusions

Structural walls are designed to resist the lateral load on the structure due to earthquake loading. With the advancement in current regression analysis and the continuing popularity of boosting algorithms. From the study of these models, it was concluded that the CATBoost gave the best prediction followed by Random Forest and XGBoost Model. The Ensemble Tree models were used to conclude that Aspect Ratio is the most important parameter affecting Shear Strength followed by Boundary Element Area to Gross Area, Length to thickness ratio. The model can be used in design offices to predict and design for large samples of shear walls and also help in taking decisions on retrofitting of existing walls. The model is flexible and new data can be added to the current database to improve the model and find more insights on factors affecting peak shear strength.

REFERENCES

[1] Thomas Paulay and MJ Nigel Priestley. *Seismic design of reinforced concrete and masonry buildings*, volume 768. Wiley New York, 1992.

[2] Ioannis D Lefas, Michael D Kotsovos, and Nicholas N Ambraseys. Behavior of reinforced concrete structural walls: strength, deformation characteristics, and failure mechanism. *Structural Journal*, 87(1):23–31, 1990.

[3] C Kerem Gulec and Andrew S Whittaker. Empirical equations for peak shear strength of low aspect ratio reinforced concrete walls. *ACI Structural Journal*, 108(1), 2011.

[4] Cevdet K Gulec, Andrew S Whittaker, and Bozidar Stojadinovic. Peak shear strength of squat reinforced concrete walls with boundary barbells or flanges. *ACI structural journal*, 106(3):368, 2009.

[5] Cevdet Kerem Gulec. *Performance-based assessment and design of squat reinforced concrete shear walls*. State University of New York at Buffalo, 2009.

[6] Panatchai Chetchotisak, Weerapong Chomchaipol, Jaruek Teerawong, and Somboon Shaingchin. Strut-and-tie model for predicting shear strength of squat shear walls under earthquake loads. *Engineering Structures*, 256:114042, 2022.

[7] Sujith Mangalathu and Jong-Su Jeon. Machine learning–based failure mode recognition of circular reinforced concrete bridge columns: Comparative study. *Journal of Structural Engineering*, 145(10):04019104, 2019.

[8] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, et al. *An introduction to statistical learning*, volume 112. Springer, 2013.

[9] Mark Fenner. *Machine learning with Python for everyone*. Addison-Wesley Professional, 2019.

[10] Andreas Mayr, Harald Binder, Olaf Gefeller, and Matthias Schmid. The evolution of boosting algorithms. *Methods of information in medicine*, 53(06):419–427, 2014.

[11] Robert E Schapire. Explaining adaboost. In *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, pages 37–52. Springer, 2013.

[12] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[13] Anna Veronika Dorogush, Vasily Ershov, and Andrey Gulin. Catboost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*, 2018.