



Machine learning prediction of fundamental time periods of regular RC bare frame buildings: A comparative study with empirical code formulations for seismic disaster risk reduction

Aniket N Bhalkikar , Jacob Alex Kollerathu*

Civil Engineering Department, CHRIST (Deemed to be University), Bangalore, India

Abstract

This study investigates the prediction of fundamental time periods of reinforced concrete (RC) bare-frame buildings using machine learning (ML) techniques as an alternative to empirical height-based code formulas. A synthetic dataset comprising 292 regular RC frame configurations was generated via parametric modelling using the ETABS API, varying key structural parameters such as building height, base width, height-to-base ratio, mass, and stiffness. Seven supervised ML regression algorithms—Linear Regression, Support Vector Regression, Decision Tree Regressor, Random Forest Regressor, Adaptive Boosting, Gradient Boosting, and XGBoost—were trained using 80% of the dataset and validated with 10-fold cross-validation. Model performance was evaluated using R^2 , RMSE, MAE, and the a_{20} index, and the results were compared with empirical formulations from IS 1893:2016 and prominent literature-based models. Ensemble boosting models exhibited superior accuracy, with AdaBoost achieving the highest predictive performance ($R^2 = 0.969$, $MAE = 0.050$ s, $a_{20} = 0.90$). SHapley Additive exPlanations (SHAP) analysis revealed mass and stiffness as the most influential parameters controlling the time period, unlike code formulations dependent solely on height. Comparative analysis demonstrated that empirical models significantly overestimate or underestimate period values with large scatter, whereas ML-based predictions closely match calculated values with minimal deviation. The findings confirm the capability of ML models, particularly AdaBoost, to improve reliability in seismic design and assessment of existing RC frames, supporting seismic vulnerability assessment and disaster risk reduction efforts. Future extensions may include neural network architectures and irregular structural configurations.

Keywords: Fundamental time period, Machine Learning (ML), Reinforced concrete frames, Ensemble boosting algorithms, ETABS API, Empirical period formula, Risk Reduction

Received: Feb 10, 2026

Revised: Apr 12, 2026

Accepted: Apr 15, 2026

Citation

Bhalkikar, A.N., Kollerathu, J.A., 2026. Machine learning prediction of fundamental time periods of regular RC bare frame buildings: A comparative study with empirical code formulations for seismic disaster risk reduction. *Int. J. Disast. Studies Climate Resil.* 2(1), 95–115.

<https://doi.org/10.64866/j.ijdsr.2026.10026>



1 Introduction

The time period of any structure is the duration that it takes to accomplish a full cycle of free vibration. It is a vital parameter in earthquake-resistant design, as it dictates a structure's response to dynamic seismic loads. The time period of a structure reflects its rigidity and mass distribution and indicates the building's flexibility or rigidity. A shorter value of period indicates a rigid structure, whereas a longer period represents a more flexible structure. The empirical equations provided in the literature to estimate the fundamental time period of RC bare-frame buildings normally assume the power-law relationship in the form:

$$T = CH^x \quad (1)$$

- T is the fundamental time period in seconds,
- H is the height in meters, and
- C and x are constants.

Seismic design codes worldwide, viz. IS 1893:2016 (IS 1893:2016, 2016), ASCE7 (ASCE7-22, 2022), and Eurocode 8 (Eurocode 8, 2004) CEN 2004 assumes uniform stiffness and mass distribution, recommending similar expressions with fixed values for C and x . For example, for RC bare frame structures, the IS 1893:2016 specifies $C = 0.075$ and $x = 0.75$.

Accurate estimation of the fundamental time period is not merely a technical design requirement; it is a critical

*Corresponding author. Email: jacob.alex@christuniversity.in (JAK), aniket.nitin@christuniversity.in (ANB)

element for effective seismic disaster design and risk reduction. An overestimated period leads to an underestimated design base shear, yielding new structures that are unsafe, structures with elevated seismic vulnerability, and improper choices for retrofitting schemes for existing buildings. Conversely, underestimation results in over-designed, uneconomical structures. Both outcomes have direct implications for building stock resilience and community-level disaster preparedness. The development of reliable, multi-parameter predictive models therefore supports rapid seismic screening of existing buildings—a key component of pre-disaster risk assessment frameworks as advocated by the Sendai Framework for Disaster Risk Reduction 2015–2030.

However, various studies, such as [Goel and Chopra \(1997\)](#), [Inel et al. \(2019\)](#), and [Velani and Ramancharla \(2023\)](#) have indicated that these empirical formulas suggested by the design codes may not fully capture the variability in the dynamic response of structures, especially with those with stiffness and geometric variations and irregularities in the vertical and horizontal directions. These studies, through numerical modelling and regression analysis, propose revised coefficients of the constants in Eq. (1), thereby serving as potential improvements in predicting the dynamic characteristics of RC bare-frame buildings.

Numerous empirical formulas have been proposed over the years to improve upon code-based expressions. Research by [Kewate and Murudi \(2018\)](#) and [Ajay Kumar and Pradeep Kumar \(2013\)](#) has proposed empirical formulas to improve the code-based expressions. These incorporate modifications based on the observed building characteristics. Despite the notable improvements of these expressions to predict the time periods of the structures, their applicability is limited when it comes to diverse structural configurations, as demonstrated by [Goel and Chopra \(1997\)](#), [Inel et al. \(2019\)](#), and [Velani and Ramancharla \(2023\)](#). This is because these empirical expressions do not fully capture the variability in the dynamic response of RC structures with vertical and horizontal irregularities in geometry and stiffness. Despite these advances, gaps remain in accurately predicting the fundamental time period, particularly regarding the influence of multiple structural parameters and the interpretability of predictive models. To address these challenges, this study introduces several novel elements:

- **Systematic Synthetic Dataset Generation:** A comprehensive synthetic database of time periods for well-proportioned RC frames using the ETABS API was developed by systematically varying key structural and geometrical parameters. This approach allows robust model training and validation by incorporating multiple structural parameters (height, mass, stiffness, base width) rather than relying solely on building height.
- **Advanced Machine Learning Application:** Seven machine learning (ML) algorithms—including ensemble methods such as AdaBoost, Gradient Boosting, and XGBoost—are employed and rigorously compared. Model Interpretability via SHAP Analysis: The use of Shapley Additive Explanations (SHAP) provides transparent, quantitative insights into fea-

ture importance, allowing engineers to understand ML predictions—a step beyond traditional ‘black-box’ models.

- **Benchmarking Against Codes and Literature:** We benchmark ML predictions against multiple code-based and literature-based formulas, demonstrating superior accuracy and reliability. Notably, our SHAP analysis reveals mass and stiffness as critical features—often neglected in code-based formulas.

An incorrect estimate of a structure’s time period results in an erroneous design force calculation, leading to unsafe, uneconomical, and non-conservative designs. Hence, the accurate estimation of the fundamental time period is a critical step toward achieving safety and economy in structural design. While seismic codes primarily use expressions based on building height, researchers have demonstrated that other structural parameters, such as material properties, also influence the time period ([Inel et al., 2019](#)). In line with this, [Li et al. \(2024\)](#), [Joseph et al. \(2024\)](#) and [Guo et al. \(2023\)](#) have adopted ML to understand the nonlinear relationships that define the dynamic responses of the building. ML techniques can model complex relationships directly from data. To date, limited studies have been reported on the development of ML models for predicting the time periods of reinforced concrete (RC) frames. Before implementing ML techniques, systematic classification and organisation of relevant data are required, followed by rigorous evaluation of each model’s prediction accuracy. In addition, the internal mechanics of the ML models are essential for understanding the predictions and the observed behavioural trends. Finally, the performance of these models should be benchmarked against standard design codes and literature to assess their reliability. This study aims to bridge the existing research gap by:

1. Compiling a comprehensive synthetic database of time periods for well-proportioned RC frames using the ETABS-API,
2. Applying various ML algorithms to predict time periods in RC frames,
3. Interpreting the model behaviour using Shapley Additive exPlanations (SHAP), and
4. Comparing ML model predictions with code-based and literature-based empirical formulations for the fundamental time period of RC frames.

This study aims to provide a transformative pathway for accurate, interpretable, and regionally adaptable time-period prediction in RC buildings through the contributions outlined above. In this regard, seven ML algorithms—Linear Regression, Support Vector Regressor, Decision Tree Regressor, Random Forest Regressor, Gradient Boosting Regressor, AdaBoost Regressor, and Extreme Gradient Boosting—are employed to train and validate the dataset. The models are trained and tested using an 80:20 split of the dataset. Furthermore, predictions from national and international design codes and research for the same experimental set are compared with the ML outputs. This study

identifies the most effective ML model for predicting time periods of RC frames and explains the underlying trends.

2 Literature Review

A fundamental parameter that governs the seismic design of RC buildings is the fundamental time period, as it influences the calculation of base shear demand and the evaluation of dynamic response. Seismic codes such as IS 1893:2016 (IS 1893:2016, 2016), ASCE 7 (ASCE7-22, 2022), and Eurocode 8 (Eurocode 8, 2004) prescribe empirical formulas in the form $T = C \cdot H^x$, where H is the building height and C, x are constants derived from regression analyses of vibration data (Badkoubeh and Massumi, 2017; Jit-teh et al., 2025; Di Sarno and Amiri, 2019). These formulas often neglect key structural features like the presence and the contributions of bracing, infill walls, slab thickness, and soil-structure interaction (SSI) (Elgohary, 2025; Balendra et al., 2003; Masi and Vona, 2010).

Numerous studies (Cinar et al., 2024; Ali et al., 2025; Elgohary, 2025; Balendra et al., 2003; Dilmaç, 2020; Gallipoli et al., 2023; Sharma et al., 2023; Sarma et al., 2022; Arlekar and Murty, 2000) have demonstrated that existing assumptions regarding the types of bracings, slab thickness, infill thickness, and floor areas lead to significant predictive errors. Regional studies from Italy and Southeast Asia, models by Masi and Vona (2010), Xiong et al. (2016), Banik et al. (2025), Goel and Chopra (1997), have shown that code-based predictions often overestimate the time period, even for damaged buildings, highlighting the need for better models.

Researchers have proposed multi-parameter empirical models that incorporate the type and length of bracing, slab thickness, infill stiffness, and floor area (Elgohary, 2025; Balendra et al., 2003; Ukwaththa et al., 2024; Cinar et al., 2024; Kewate and Murudi, 2018; Giri et al., 2019; Mohamed et al., 2019). Other models have included regression models to expand further the feature set to include building width, number of bays, and infill condition (Jose and Mishra, 2021). Recent work has seen the adoption of Rayleigh's approach to determine the the time periods that is observed to outperform standard code formulas (Panzer et al., 2013; Shendkar et al., 2021; Hadzima-Nyarko et al., 2015; Bioud et al., 2023; Zhang et al., 2023; Sarma et al., 2022; Ukwaththa et al., 2024; Ma and Wang, 2023; Nguyen et al., 2024).

ML-based models have also shown promise in post-earthquake assessment, with damage-inclusive models accurately predicting time periods for both damaged and undamaged buildings (Bioud et al., 2023; Shan et al., 2024). SSI-inclusive models using LSTM, GRU, and 1D-CNN architectures have outperformed fixed-base assumptions, particularly for buildings on soft soils and pile foundations (Nguyen et al., 2025; Rahman et al., 2025; Calitz et al., 2023). These advances have led to the development of practical tools such as mobile apps and web dashboards for real-time period estimation (Kewate and Murudi, 2018; Ditommaso et al., 2013; Hadzima-Nyarko et al., 2015).

Despite these advancements, challenges remain in integrating damage states, SSI effects, and regional construc-

tion practices into standardised design frameworks. Future research directs expanding datasets to include irregular buildings and post-earthquake damage data, refining the selection of important features that govern time period calculation, and collaborating with code committees to validate and implement ML models (Bioud et al., 2023; Ahmad et al., 2025; Narkhede, 2024).

3 Need for Research

An overestimation of the time period of a structure results in a reduced estimate of the design force, leading to unsafe designs, while underestimation leads to overly conservative, uneconomical solutions. Thus, a correct estimate of the natural time period is not only a technical requirement but also a critical step toward achieving safe, cost-effective structural design. This study develops a synthetic dataset from ETABS-API to train and validate ML regression algorithms for predicting the fundamental time periods of RC frames. To estimate the reliability of ML predictions, seven algorithms are compared with four existing methods from the literature. These comparisons assess the consistency and precision of ML-based predictions relative to conventional code-based approaches. A significant deviation between predicted time periods and code-based values may indicate potential concerns about seismic performance, warranting further investigation from a safety and serviceability perspective.

4 Machine Learning Algorithms

In this paper, bagging and boosting (Fig. 1) are used to classify machine learning regression algorithms. The difference between the two lies in how they handle the subset. While bagging algorithms use parallel processing of the data, boosting algorithms use sequential processing. Bagging algorithms can prevent overfitting by reducing the variance. Boosting algorithms can improve prediction accuracy by reducing bias.

The choice of the appropriate ML model depends significantly on the database and its training data characteristics. Neural networks have not been employed in this paper due to the limited data available. The ML models are evaluated on four features (x_i), namely: the height of the frame, h (m); the width of the frame, b (m); the mass of the frame, m (kg); and the stiffness of the frame, K (kN/m). The target variable, $y(x)$, is the fundamental time period of the frame, T (s).

4.1 Linear Regression (LR)

It is a fundamental regression method to model a linear relationship between a dependent variable and one or more independent variables, typically represented as:

$$y_i = w_i x_i + b \quad (2)$$

where y_i is the target variable (response), x_i is the independent variable, b is the intercept, and w_i is the coefficient for each predictor/feature.

The values of the intercepts to determine the best fit of predictions calculated from (2) are obtained by evaluating

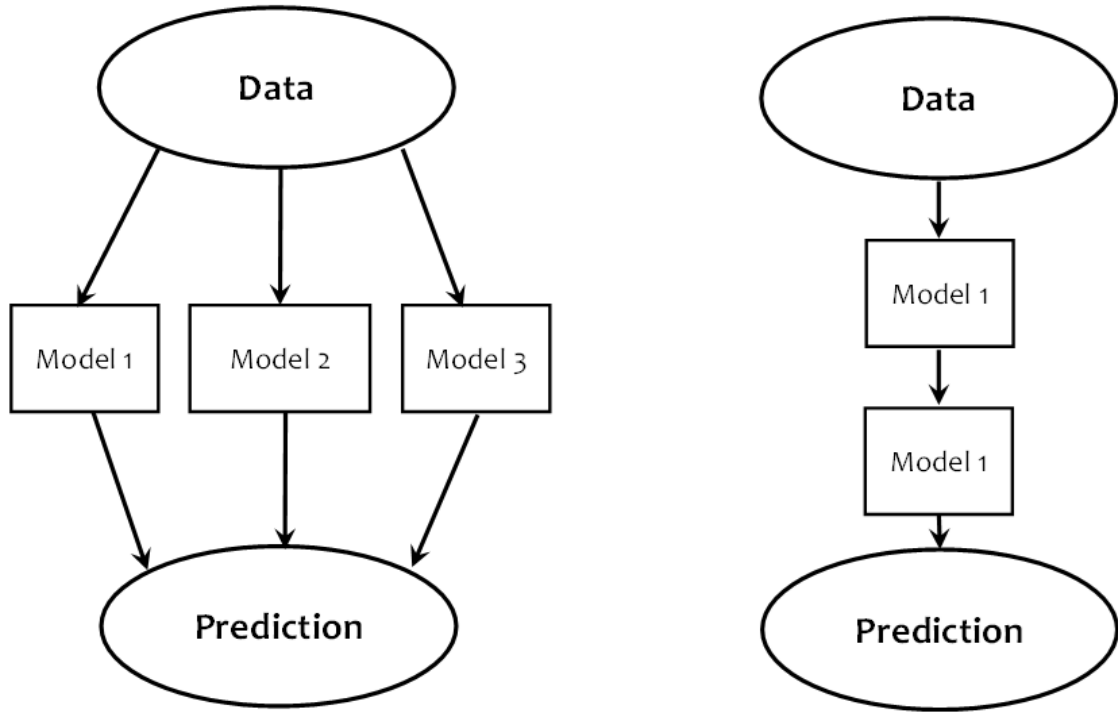


Fig. 1. Difference between bagging and boosting.

the least square error expression and then differentiating the expression with respect to the intercept and the slope. The error function is typically obtained as:

$$e = y_{\text{act}} - y_{\text{pred}} \quad (3)$$

where y_{pred} is the predicted value of the target variable, and y_{act} is the actual value of the target variable. The mean square error of the error function is then obtained as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i^a - y_i^p)^2 \quad (4)$$

where y_i^a is the actual value and y_i^p is the predicted value. To minimise the errors and to determine the accurate values of the intercept and the coefficients, Eq. (4) is differentiated with respect to w and x_i .

4.2 Support Vector Regression (SVR)

SVR is a regression algorithm based on the Support Vector Machines (SVMs). It is an extension of the pattern classifier problem with multidimensional input parameters (Cortes and Vapnik, 1995) and is currently used for prediction, classification and regression. Instead of minimising the error as in linear regression, SVR tries the best fit within a tolerance (ε). It focuses on points within the margin of tolerance, ignoring errors less than ε and peeling those larger than the margin. SVR is represented as:

$$y = \sum_{i=1}^n w_i \phi_i + b \quad (5)$$

where w , ϕ_i , and b are the weight vectors, the non-linear kernel function and the bias. The loss function $L(y)$ is represented by:

$$L(y)f(x) = \begin{cases} 0, & |y - f(x)| \leq \varepsilon \\ |y - f(x)|, & \text{otherwise} \end{cases} \quad (6)$$

4.3 Decision Tree Regressor (DTR)

The DTR model is a non-parametric supervised learning model that uses conditional statements to predict continuous numerical values. This method was introduced by Breiman et al. (1984) and is currently used for classification, prediction, and regression. It is important to note that a single Decision Tree is not a bagging algorithm; bagging-based ensemble learning (e.g., Random Forest) aggregates predictions from multiple trees. DTR works by identifying the root node and splitting the data into subsets to reduce the mean-squared error. The loss function is represented in Eq. (7):

$$l(y_a, y_p) = |y_a - y_p| \quad (7)$$

where y_a and y_p are the actual and predicted values, respectively. The error function is represented as:

$$\epsilon = \frac{1}{N} \sum [l(y_n), f(x_n)] \quad (8)$$

4.4 Random Forest Regressor (RFR)

The RFR algorithm was developed by Breiman (1996). It is a random combination of a few features and bootstrap aggregation (bagging). An RFR model is a supervised ensemble learning method that provides superior performance in regression and classification. Ensemble models are learning algorithms that develop a set of classifiers/regressors by combining weighted individual predictions (Dietterich, 2000). An RFR model is a bagging ensemble of decision tree models, consisting of root nodes and child nodes/leaf

Table 1. Summary of Dataset Configuration Parameters.

Parameter	Range / Value
Number of storeys	2 to 18 (storey height = 3 m, uniform)
Number of bays (each direction)	2 to 6
Bay width	3.0 m to 7.5 m
Column cross-section	0.35 m × 0.35 m to 0.70 m × 0.70 m (square)
Beam cross-section	300 mm × 500 mm (fixed across all models)
Concrete grade	M25 (fck = 25 N/mm ²)
Slab thickness	125 mm
Boundary conditions	Fixed base
Element type	3D frame elements with rigid diaphragm assumption
Analysis procedure	Linear eigenvalue (modal) analysis via ETABS API
Plan regularity	Regular (symmetric in both principal directions)
Elevation regularity	Regular (uniform storey heights, no setbacks)
Infill walls	Not modelled (bare frame only)
Soil-structure interaction	Not considered (fixed base assumed)

nodes. The root node is identified using the Gini Index (GI).

$$GI = 1 - \sum_{i=1}^n P_i^2 = 1 - \sum_{i=1}^n (P_A^2 - P_B^2) \quad (9)$$

where, P_A is Probability of a positive outcome, and P_B is probability of a negative outcome. The RFR provides advantages such as handling nonlinear data, reducing overfitting, and handling missing data.

4.5 Adaptive Boosting Regressor (AdaBoost)

Boosting algorithms are a widely used machine learning algorithm, proposed by [Schapire \(2013\)](#). AdaBoost combines a weak and a strong classifier. AdaBoost was introduced to combine the outputs of weak learners gained during training to create a strong regressor ([Chengsheng et al., 2017](#)).

4.6 Gradient Boosting Regressor

The Gradient Boosting Regressor is a stable, highly accurate algorithm that uses numerous base estimators. Distinct DTR models are generated using Boosting. The predicted models, aided by residuals, are optimised by a sequentially arranged tree based on errors calculated from previous steps. The loss function (l) is optimised using a negative gradient descent.

$$l = \frac{1}{N} \sum_{i=1}^n (y_a - y_p)^2 \quad (10)$$

where, y_a : Actual value and y_p : Predicted value. The final output value is the mean of all predictions and is evaluated as shown below:

$$\gamma_m = \operatorname{argmin} \sum_{i=1}^n [l, y_i, f_{m-1}(x_i) + \gamma h_m(x_i)] \quad (11)$$

where γ_m is the final prediction, $h(x_i)$ is the decision made on residuals, m is the number of decision trees, and γ is the final output.

4.7 Extreme gradient boosting (XGBoost) Regressor

The XGBoost model is a tree-based gradient boosting ensemble model, developed by [Chen and Guestrin \(2016\)](#). It is one of the most widely used and popular models to solve well-structured tabular data. The gradient boosting architecture uses various complementary functions to determine the result, as shown by Eq. (12).

$$y_i = y_i^0 + \sum_{i=1}^n f_k U(i) \quad (12)$$

where, y_i is the predicted output, (i) is the parameter vector, n denotes the number of estimators, and y_i^0 denotes the primary hypothesis. The objective function in Eq. (13), also called the loss function, plays an important role in minimising error during model training.

$$\operatorname{obj}(\theta) = \sum_{i=1}^n l(y_i - \bar{y}_i) + \sum_{k=1}^K \Omega(fk) \quad (13)$$

5 Description of Dataset

The study generated 292 bare-frame RC building models through systematic parameter variation. All structural configurations were developed using ETABS API and stored in an Excel workbook for machine learning integration. The buildings are regular in plan and elevation, with uniform storey heights and symmetric bay layouts, ensuring that the fundamental mode is the dominant mode of vibration (modal mass participation ratio typically exceeding 80%). The complete parametric description of the models is provided in [Table 1](#).

The column sizes were systematically determined according to both building height and bay configurations. Standard square columns ranged from 0.35 to 0.7 m, with increments of 0.05 m between sizes. For low-rise buildings under 30 m in height, column dimensions typically ranged between 0.35 and 0.5 m, while mid-rise structures exceeding 30 m utilised larger columns, measuring 0.55 to 0.7 m. Additionally, the column size was adjusted proportionally to bay lengths, with an increase of 0.05 m in column dimension for

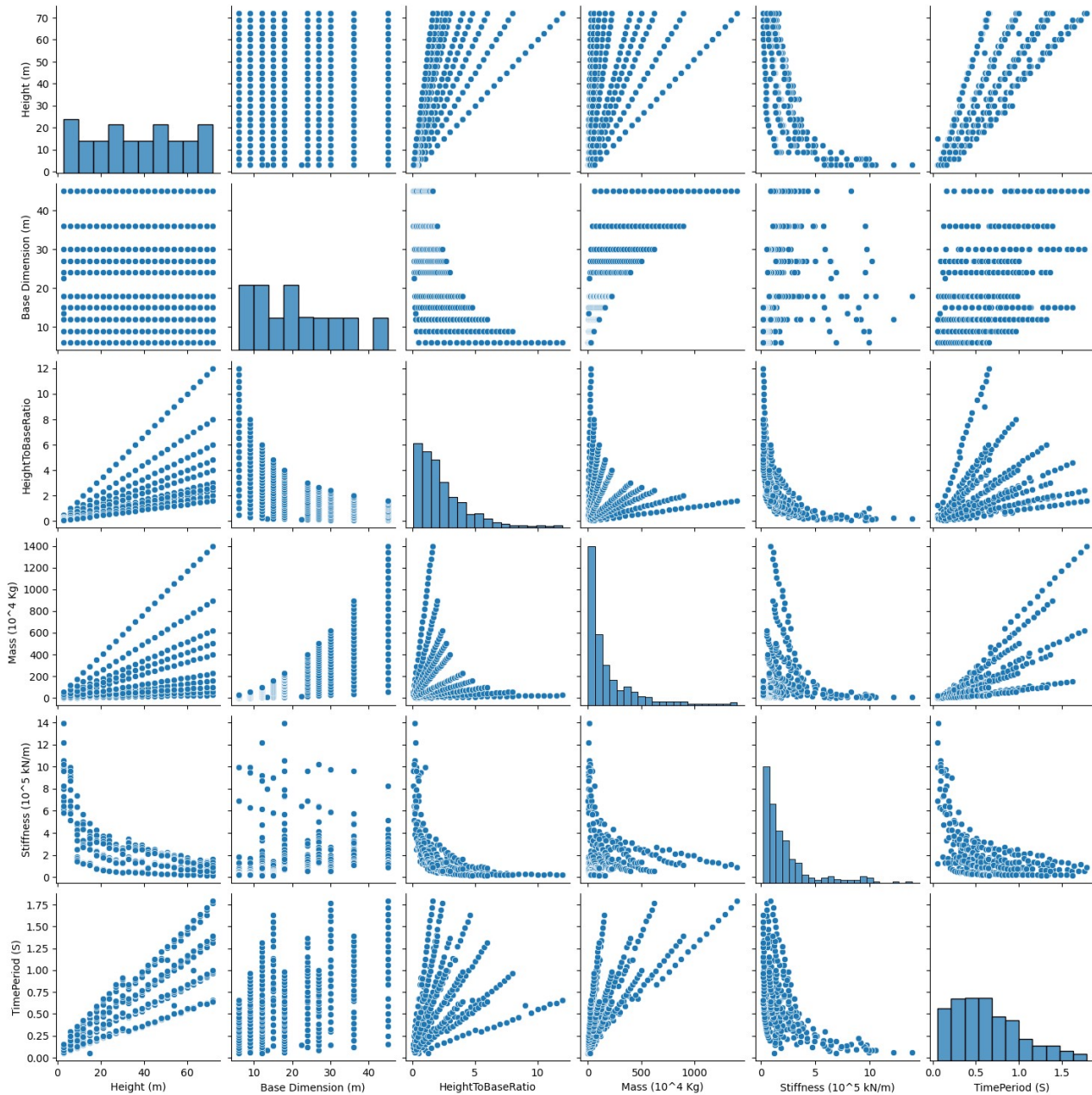


Fig. 2. Relationship of input variables with time period.

every 3 m expansion in span length. This sizing methodology ensured appropriate structural proportions across various building configurations. The datasets used in this study are available from the corresponding author on reasonable request, subject to approval for academic or research purposes.

5.1 Statistical Description and Correlation Analysis of the Dataset

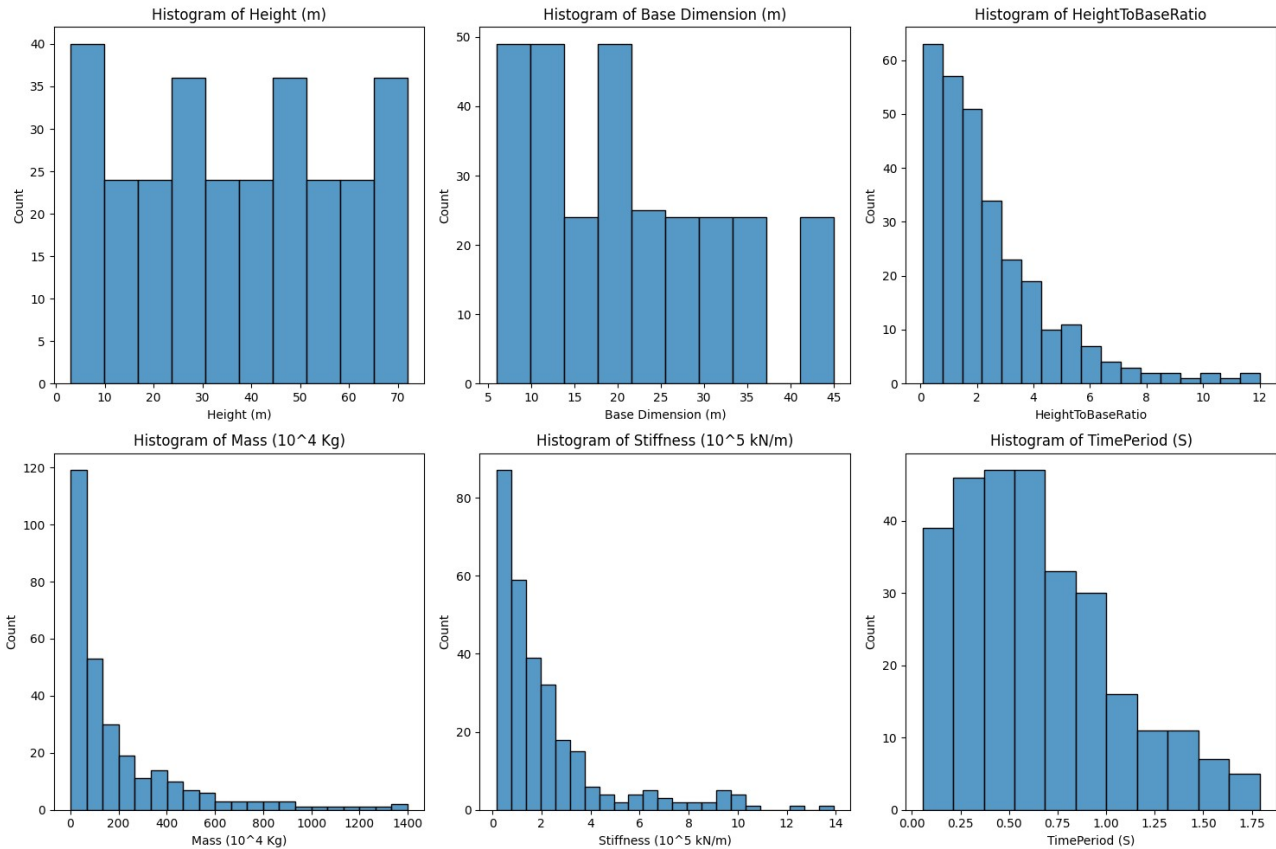
Table 1 represents the statistical representation of the dataset and the relationship between input variables and time period is illustrated in Fig. 2. Table 2 summarises the maximum and minimum values, standard deviation (SD), quartiles, and mean. The statistical analysis illustrates that the large range covered by the dataset, and SD represents the dataset based on its distribution along the mean values.

The analysis of the ML models was carried out based on five input parameters viz. height of the frame (m), width of the frame at the base (m), and height-to-base ratio, mass of the frame and stiffness, and time period as output parameters.

As per Fig. 3, the range of the time periods is 0.06–1.79 s. The figure also shows that the parameters used in the database cover buildings with a wide range of geometrical features, mass and stiffness, which proves its reliability. Hence, the proposed model can be used to predict the time periods of a wide variety of geometrical RC frames. Using correlation analysis, input variables that are not related to the RC frame time periods should be excluded from the dataset. The most commonly used correlation coefficient is the Pearson correlation. The Pearson correlation coefficient ρ_{xy} , described in Eq. (14), defined as the ratio of the covariance of two variables to the product of their standard

Table 2. Statistical description of dataset.

Statistic	Height (m)	Base Dimension (m)	Height to Base Ratio	Mass (10 ⁴ kg)	Stiffness (10 ⁵ kN/m)	Time Period (s)
Std	21.05	11.20	2.18	260.22	2.43	0.40
Min	3.00	6.00	0.07	1.04	0.17	0.06
Mean	37.03	20.93	2.42	201.43	2.19	0.64
Max	72.00	45.00	12.00	1399.68	13.91	1.79
75%	54.00	27.00	3.27	259.72	2.62	0.89
50% (median)	36.00	18.00	1.79	95.39	1.37	0.58
25%	18.00	12.00	0.90	34.54	0.62	0.33

**Fig. 3.** Frequency distribution of the variables.

deviations.

$$\rho_{xy} = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (14)$$

The value of the Pearson coefficient ranges between (-1, 1). A value of ρ_{xy} close to 1 indicates that the parameters are directly related and have a high effect on the output parameter, while $\rho_{xy} = -1$ denotes an inverse correlation. Also, if $\rho_{xy} = 0$, it represents parameters are independent.

The analysis of the parameter-period relationships in Fig. 4 reveals several key patterns. It is the height that demonstrates the strongest influence on fundamental period with a 0.78 correlation. Mass follows with a 0.70 correlation, though the relationship becomes more pronounced for heavier structures. Stiffness exhibits the expected inverse relationship at -0.55 , where increased stiffness consistently reduces period duration. Base width shows a more moder-

ate 0.38 correlation with greater data dispersion compared to other parameters. These results confirm that while all four parameters contribute to period determination, their relative impacts follow a clear hierarchy from strongest to weakest: height, mass, stiffness, and base width. The visual patterns quantitatively validate fundamental structural dynamics principles, particularly the dominant role of building height in the time period prediction.

The height-to-base ratio, being a derived quantity from height and base width, introduces potential multicollinearity into the feature set. A Variance Inflation Factor (VIF) analysis was conducted, yielding a VIF of 5.8 for the height-to-base ratio. While this value is elevated, it remains below the commonly used threshold of 10 that typically indicates problematic multicollinearity. Moreover, tree-based ensemble methods (Random Forest, AdaBoost, Gradient Boosting, XGBoost) are inherently robust to correlated features because they operate on random feature subsets at each

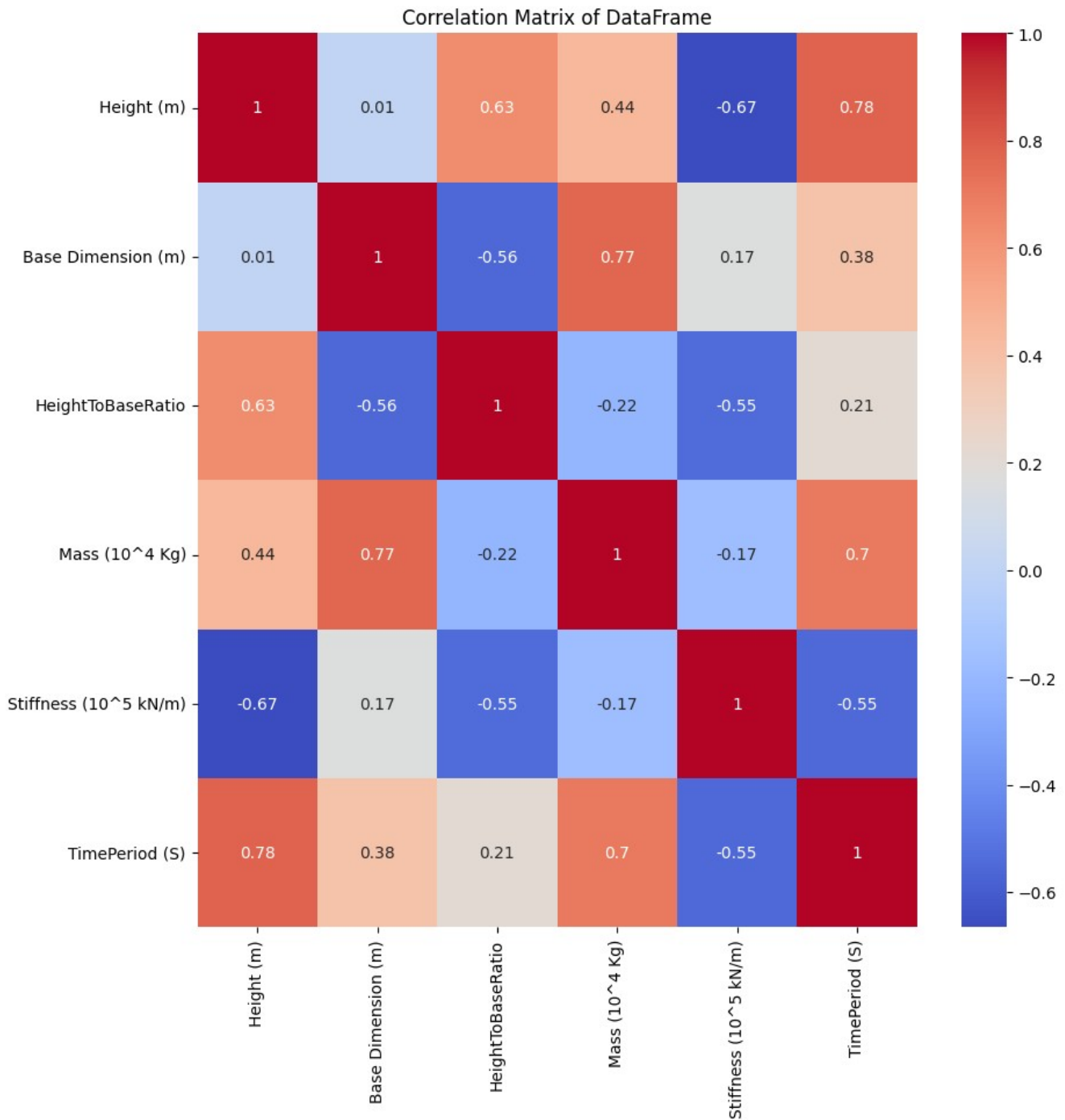


Fig. 4. Correlation relationships between the variables.

split. The height-to-base ratio was retained as it encodes the slenderness of the structure—a physically meaningful parameter for dynamic behaviour that cannot be fully captured by height and base width individually.

5.2 Data Normalization

The effectiveness of the model depends on the range of the values of the variables. From Table 1, one can observe that while the mass of the frame varies from 1.04 kg to 1400 kg, the time period varies from 0.06–1.79 s. To avoid this variation in the magnitudes, data normalization may be used wherein the range of all the variables can be brought

within the same scale. This is achieved by using the min-max function which is represented by Eq. (15).

$$x_n = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (15)$$

where x_n is the normalised data, $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of each input variable, respectively, and x is the original dataset to be rescaled. A major perk of data normalisation is speeding up computation and improving the accuracy and stability of the prediction model.

Table 3a. 10-Fold Cross-Validation Results (Mean $\pm$ Standard Deviation).

Model	Mean R^2	Std R^2	Mean RMSE (s)	Std RMSE (s)
Linear Regression	0.70	0.06	0.218	0.031
SVR	0.75	0.05	0.194	0.028
Decision Tree	0.93	0.03	0.101	0.018
Random Forest	0.95	0.02	0.086	0.014
AdaBoost	0.96	0.02	0.071	0.012
Gradient Boosting	0.95	0.02	0.079	0.013
XGBoost	0.96	0.01	0.074	0.011

6 K-Fold Validation

The K-fold validation is a method used to validate the adopted ML algorithms. K represents the number of folds that the adopted dataset is divided into for training and testing purposes. The dataset adopted was randomly separated into 10 divisions, and one of the test datasets for the experiment was taken (Ron, 1995). K-fold cross-validation is used to enhance the performance of ML models by avoiding the neglect of essential datasets. In this study, the cross-validation methodology adopted split the data into 80% training and 20% testing, respectively.

The dataset of 292 samples, while modest in absolute terms, is appropriate for a five-feature regression problem when combined with 10-fold cross-validation. Prior ML studies on similar structural engineering prediction tasks have employed comparable sample sizes (e.g., Bioud et al., 2023; Ukwaththa et al., 2024).

The output from different models is evaluated using the root mean square error (RMSE) (represented in Eq. (16)–(18)). These indicators offer perspectives on the model performance, capturing both absolute and relative errors as well as the goodness of fit.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2 \quad (16)$$

where n is the number of observations. As the squaring of errors leads to penalising larger deviations heavily, MSE is sensitive to outliers. A lower MSE indicates higher prediction accuracy. It is useful to eliminate large errors, as it disproportionately weights them.

$$RMSE = \sqrt{MSE} \quad (17)$$

RMSE provides an error metric in the same units as the predicted variable, making it more interpretable than MSE. It maintains the same sensitivity to large errors while offering a more intuitive understanding of typical prediction deviation.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (18)$$

Mean Absolute Error (MAE) measures the average of the absolute differences between predicted and actual values. Unlike MSE and RMSE, MAE treats all errors equally, without squaring. It is thus more robust to outliers and provides a straightforward measure of the average magnitude of error, making it ideal for contexts where all deviations are considered equally undesirable.

Coefficient of determination (R^2) evaluates the proportion of variance in the dependent variable that is explained by the independent variables.

$$R^2 = 1 - \frac{\sum (y_i - y')^2}{\sum (y_i - \bar{y})^2} \quad (19)$$

where y' is the mean of actual values. The R^2 value ranges from 0 to 1, with 1 indicating perfect prediction and 0 indicating that the model performs no better than simply predicting the mean. A higher R^2 reflects better explanatory power of the model, making it a key indicator of the overall goodness of fit. Together, these performance metrics provide a comprehensive assessment of model accuracy, capturing both the magnitude and variability of prediction errors. Ideally, a model is regarded as accurate when its R^2 value is more than 0.8. Both the MSE and RMSE depict the level of discrepancy between the tested and predicted results. The accuracy of the model's predictions increases as the MSE and RMSE values reduce. MAE represents the actual difference between the tested values and predicted values, and a superior performance of the ML model is when the MAE values are closer to 0.

The fold-wise statistics in Table 3a confirm model stability: ensemble boosting models show low standard deviations in R^2 ($\sigma < 0.02$) and RMSE ($\sigma < 0.013$ s) across folds, indicating that performance is not an artefact of a particular train-test split. Linear Regression and SVR show slightly higher variance across folds, consistent with their lower overall performance. Hyperparameters are adopted to control the behaviour of the model and optimise the performance. They minimise the predefined loss function and improve the performance of the model to produce better results with fewer errors. In this study, hyper-parameter tuning was adopted for the AdaBoost, Gradient Boost and the XGBoost.

7 Results and Discussion

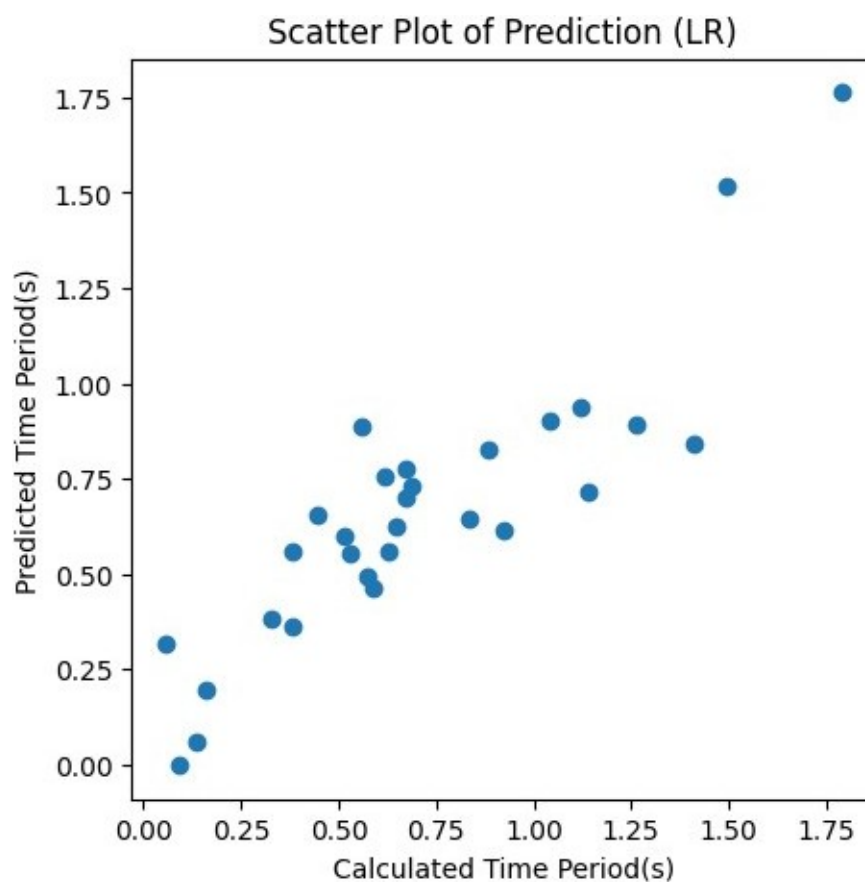
The performance of the seven adopted machine learning regression models was evaluated based on key statistical indicators including R^2 , RMSE, MAE, and the a20 index. The comparative assessment demonstrates a clear improvement in prediction accuracy from simple linear models to advanced ensemble techniques. Ensemble boosting models such as AdaBoost, Gradient Boosting, and XGBoost exhibited the highest reliability with significantly reduced error margins and minimal scatter. The results highlight the strong predictive capability of boosting algorithms for es-

Table 3b. Hyperparameter Tuning Summary.

Model	Optimal Hyperparameters (Random Search → Grid Search refinement)
AdaBoost	n_estimators = 200; learning_rate = 0.05; base_estimator = DecisionTreeRegressor(max_depth = 4)
Gradient Boosting	n_estimators = 300; learning_rate = 0.05; max_depth = 4; subsample = 0.8; min_samples_split = 4
XGBoost	n_estimators = 300; learning_rate = 0.05; max_depth = 5; subsample = 0.8; colsample_bytree = 0.8; reg_lambda = 1.0

Table 3c. Training vs. Test Performance (Overfitting Assessment).

Model	Train R^2	Test R^2	Train RMSE (s)	Test RMSE (s)
Linear Regression	0.73	0.72	0.210	0.218
SVR	0.78	0.77	0.188	0.194
Decision Tree	0.998	0.946	0.018	0.097
Random Forest	0.985	0.960	0.049	0.082
AdaBoost	0.984	0.969	0.052	0.071
Gradient Boosting	0.981	0.961	0.055	0.079
XGBoost	0.983	0.964	0.051	0.074

**Fig. 5.** Predicted by LR vs. Calculated values of time periods.

timating fundamental time periods of RC frames, notably outperforming theoretical empirical models from existing literature and seismic design codes.

To assess overfitting, training and test performance metrics are compared in Table 3c. The Decision Tree model shows the largest train-test R^2 gap (train $R^2 = 0.998$ vs test $R^2 = 0.946$), indicating mild overfitting. Boosting models show substantially smaller gaps (e.g., AdaBoost: train $R^2 = 0.984$ vs test $R^2 = 0.969$), demonstrating that regularisation through ensemble averaging and hyperparameter

tuning effectively controls overfitting for the dataset size of 292 samples (Table 3b).

7.1 Linear Regression

The comparison of actual and predicted values with the linear regression model is shown in Fig. 5. The linear regression model exhibits a moderate accuracy in prediction with a R^2 score of 0.72. The MAE and RMSE were 0.145 and 0.218 s, respectively, indicating a moderate error distribution with an MSE of 0.048 s^2 .

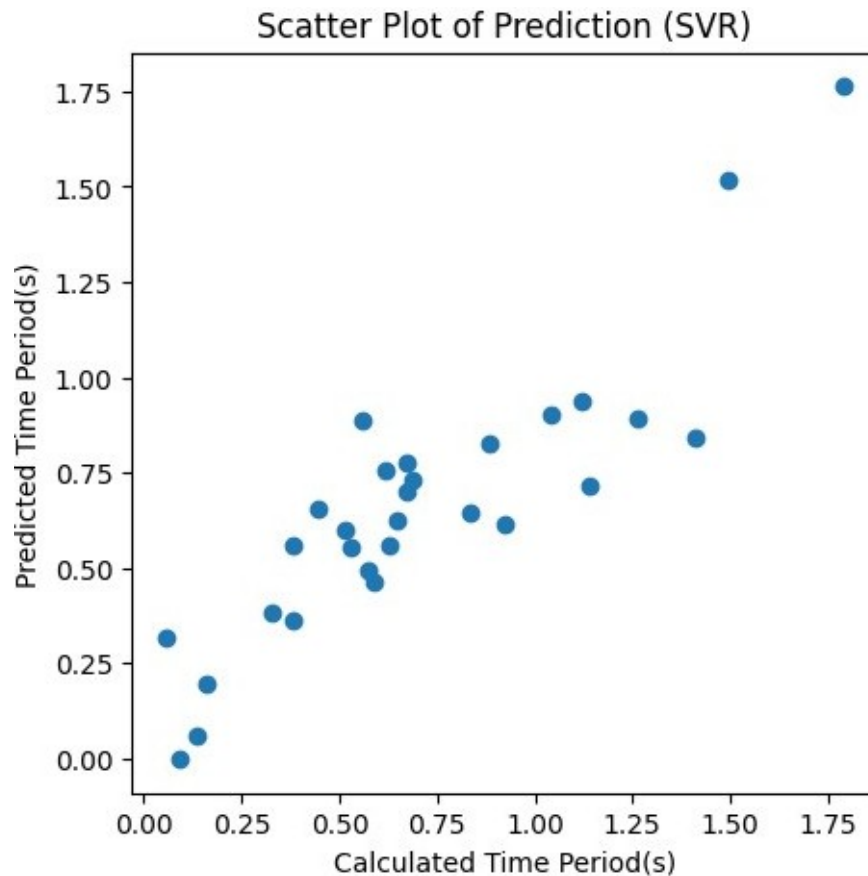


Fig. 6. Predicted by SVR vs. Calculated values of time periods.

7.2 Support Vector Regression

The comparison of actual and predicted values with the SVR model is shown in Fig. 6. In comparison with the linear regression model, the SVR model exhibits a moderate accuracy in prediction with an R^2 score of 0.77. The MAE and RMSE were obtained as 0.1404 s and 0.188 s, with an MSE of 0.035 s, showing a moderate error distribution. The mild improvement in the accuracy is attributed to the hyper-parameter tuning technique, such as random search, grid search and Bayesian Optimization. The hyperparameters improve the performance of the model and improve the results by minimising the loss function.

7.3 Decision Tree Regression

The comparison of actual and predicted values with the decision tree model is shown in Fig. 7. The decision tree regression model exhibits a high accuracy in prediction with an R^2 score of 0.946. The MAE and RMSE were obtained as 0.072 s and 0.097 s, showing a comparatively lower error distribution, corresponding to an MSE of 0.0094 s².

7.4 Random Forest Regression (RFR)

The comparison of actual and predicted values with the RFR model is shown in Fig. 8. The decision tree regression model exhibits a higher accuracy in prediction with an R^2 score of 0.962. The MAE and RMSE were 0.061 and 0.082 s, respectively, with an MSE of 0.0067 s², indicating a comparatively lower error distribution.

7.5 Adaptive Boost Regression (AdaBoost)

Adaptive Boosting (AdaBoost) is an ensemble learning method that combines several weak classifiers to create a strong classifier. At every iteration, a weak learner is trained to minimize the classification error, weighted on statistical errors from the previous iterations. The contribution to the final classifier is determined through its own accuracy across all rounds. The final prediction is set by a weighted majority vote in classification, and by a weighted average in regression. The comparison of actual and predicted values with the ensemble random forest regression model is shown in Fig. 9. The AdaBoost regression model exhibits a higher accuracy in prediction with an R^2 score of 0.969. The MAE and RMSE were 0.05 and 0.071 s, respectively, with an MSE of 0.005 s², indicating a lower error distribution.

7.6 Gradient Boost Regression (Grad Boost)

The comparison of actual and predicted values with the Gradient Boost regression model is shown in Fig. 10. The AdaBoost regression model exhibits a higher accuracy in prediction with an R^2 score of 0.9609. The MAE and RMSE were 0.06 and 0.0787 s, respectively, indicating a lower error distribution. The MSE reported is 0.0062 s².

7.7 XGBoost Regression (XGB)

Extreme Gradient Boosting (XGBoost) builds an ensemble of decision trees sequentially, where each new tree is

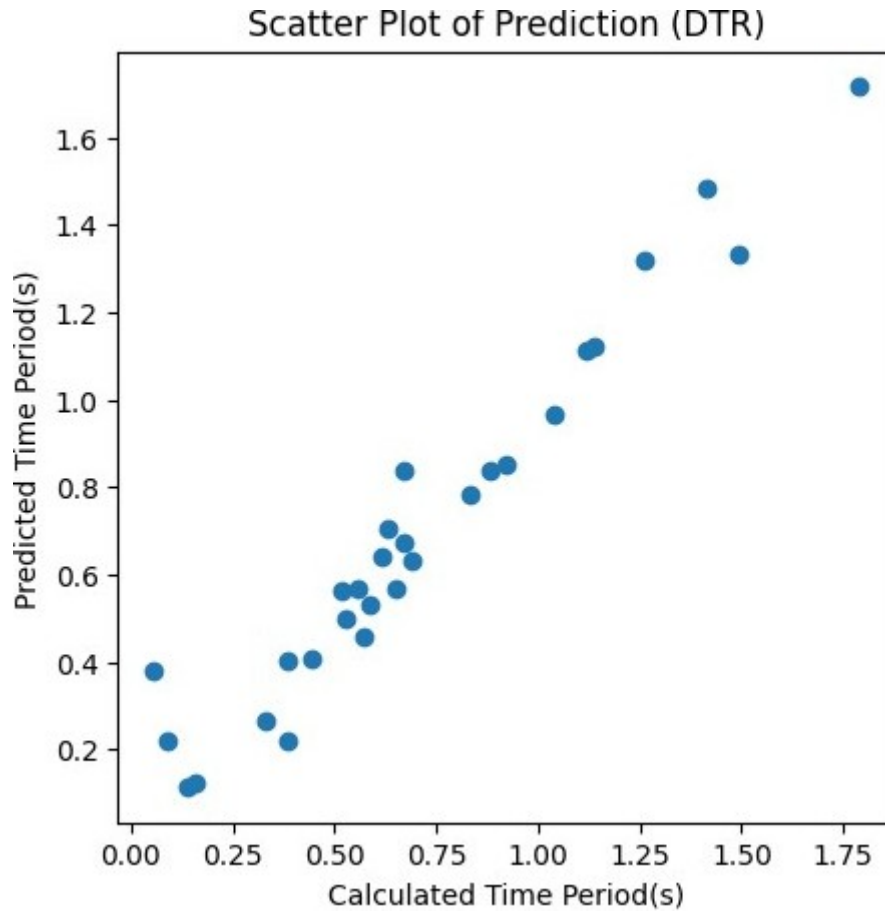


Fig. 7. Predicted by DTR vs. Calculated values of time periods.

trained to correct the residual errors of the previous trees by minimizing a regularized objective function. The comparison of actual and predicted values with the XGBoost regression model is shown in Fig. 11. The XGBoost regression model exhibits a higher accuracy in prediction with an R^2 score of 0.964. The MAE and RMSE were obtained as 0.06 s and 0.0062 s, showing a comparatively lower error distribution. The MSE is calculated as 0.006 s².

7.8 a_{20} Index

It is difficult to evaluate and compare different ML models from different slope values and varied values of R^2 . Hence, the a_{20} index is recommended for the reliability assessment of the created ML techniques. The a_{20} index, defined as the ratio of the number of samples with a prediction error of less than 20 per cent to the total number of samples, is obtained as given in Eq. (20):

$$a_{20} = \frac{m_{20}}{M} \quad (20)$$

where, M is total number of samples in the dataset and m_{20} is number of samples having a predicted value/actual value ratio between 0.8–1.2.

The perfect predictive model is characterized by a high values of R^2 , and a value of a_{20} index close to 1. The a_{20} -index of AdaBoost, Grad Boost and XGBoost was the best among the studied ML algorithms, as shown in Table 4. A similar trend is also obtained for the variation of the R^2

score (Fig. 12).

Table 4. Model evaluation using R^2 and a_{20} for the ML models.

Model	R^2 Score	a_{20} Index
Linear Regression	0.72	0.52
SVR	0.77	0.55
DTR	0.946	0.76
RFR	0.96	0.86
AdaBoost	0.969	0.90
GradBoost	0.9609	0.90
XGBoost	0.964	0.90

8 Analysis of ML models using SHAP

To understand the conclusions of complex ML models and to be able to interpret the behaviour, Shapely Additive Explanations (SHAP) are used. SHAP algorithms calculate the contribution of each feature to a prediction and has been extensively used in the applications of ML in structural engineering. The SHAP values for any application with n features is obtained as:

$$f(x) = \phi_0 + \sum_{i=1}^n \phi_i x'_i \quad (21)$$

For any feature, $x = h_x(x')$, where h_x defines the relationship between x , the dependent variables and x' the

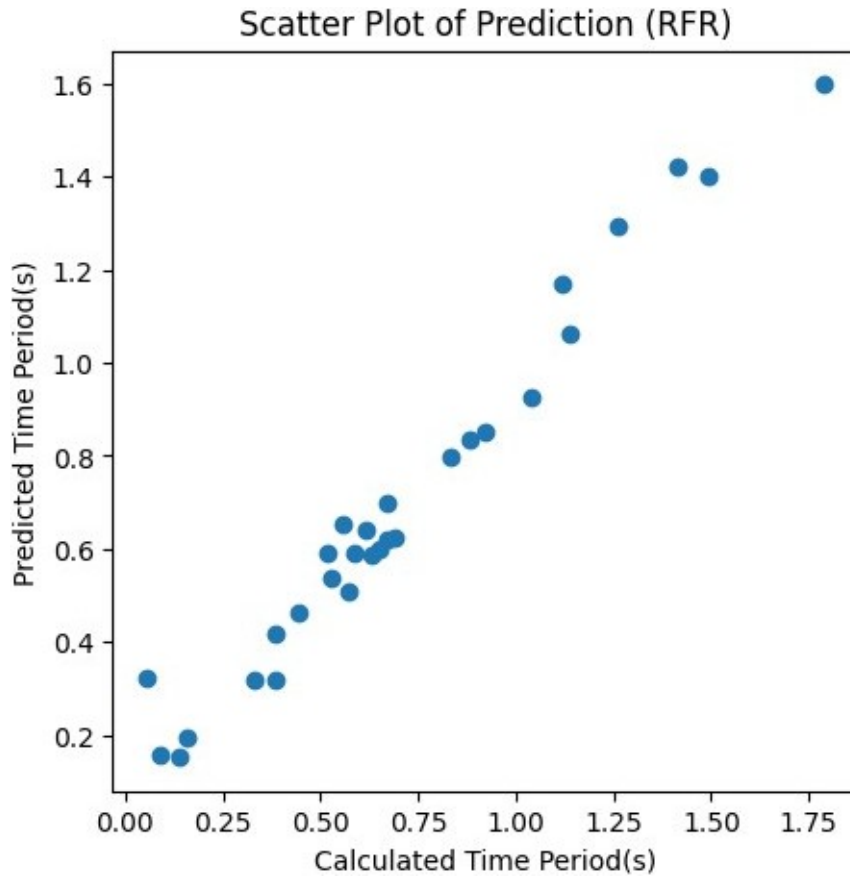


Fig. 8. Predicted by RFR vs. Calculated values of time periods.

independent variables, and $\emptyset_i$ represents the value when all the inputs are absent. By accounting for three parameters viz. local accuracy, missing values and consistency, the solution to Eq. (21) is obtained as:

$$\emptyset_i(f, x) = \frac{\sum [z']! (M - [z'] - 1)!}{M!} (f_x(z') - (f_x(z') \setminus i)) \quad (22)$$

where, $|z'|$ represents the non-zero entities in z' and $z' \subseteq x'$. Additionally, $(z') = (h(z')) = E[f(z)]zS$. S is the set of non-zero indices in z' . These values are known as SHAP values and can determine whether an input variable has a positive, negative or no contribution towards each prediction.

The Shapley values of the AdaBoost regressor are depicted using a bee swarm plot and is represented in Fig. 13. In this plot, the contribution of each feature is color-coded, with red indicating high contribution or impact on the prediction, and blue indicating low contribution. From Fig. 13 the following is interpreted:

1. *Stiffness (10^5 kN/m):*

High stiffness values (pink dots) tend to slightly decrease the output (negative SHAP values). Low stiffness values (blue dots) can increase the prediction.

2. *Mass (10^4 Kg):*

Higher mass (pink) is associated with

higher SHAP values (positive contribution). Lower mass (blue) pulls predictions lower.

3. *Height (m):*

Lower heights (blue) slightly decrease the output. Higher heights (pink) generally increase the output.

4. *Base Dimension (m):*

Higher base dimensions (pink) push the prediction higher. Lower base dimensions (blue) have small or negative contributions.

The SHAP summary plot shows that mass and stiffness are the most influential features in the model. The SHAP analysis of the AdaBoost model revealed that mass and stiffness are the two most influential features driving time-period predictions, with higher mass and larger base widths pushing predicted periods upward and greater stiffness pulling them downward; height plays a secondary yet still positive role, and low values of these parameters produce correspondingly opposite effects. By quantifying each feature's contribution to individual predictions, the SHAP summary plot not only confirms the physical logic behind the model but also enhances trust and interpretability, showing exactly how and why the ensemble regressor arrives at its estimates.

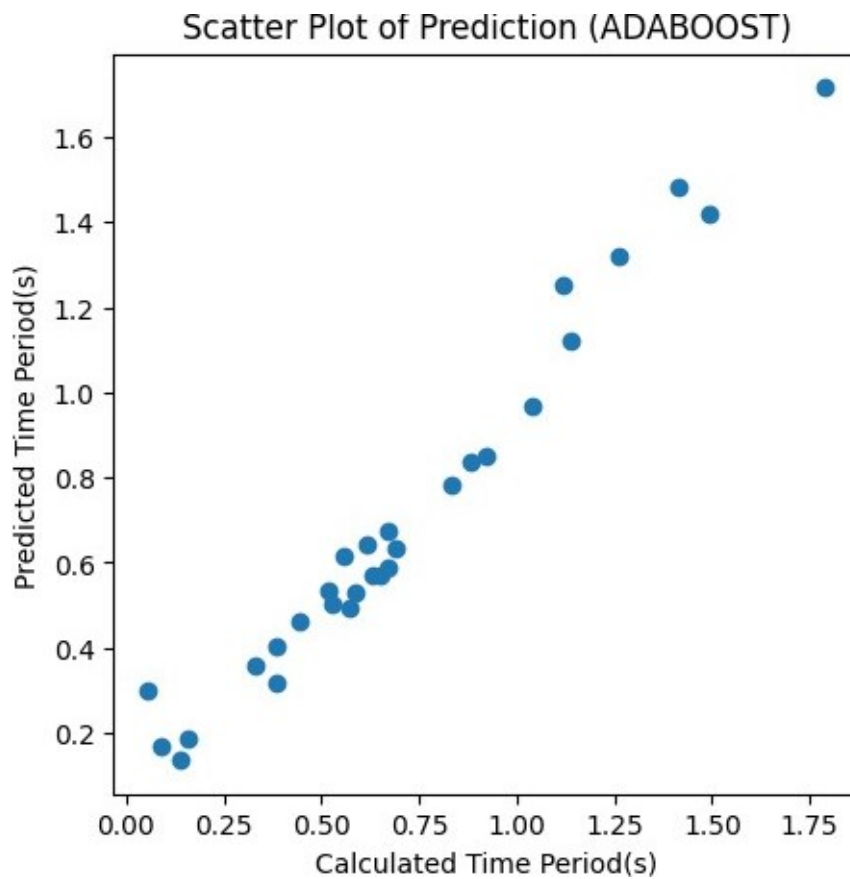


Fig. 9. Predicted by AdaBoost vs. Calculated Values of time periods.

9 Prediction of Time Periods using Standard Codes of Design and Existing Literature

The performance of the ML models was evaluated by comparing it against prominent empirical formulations proposed in the literature, such as (Velani and Ramancharla, 2023) (Inel et al., 2019) (Goel and Chopra, 1997), (IS 1893:2016, 2016). Key evaluation metrics—coefficient of determination (R^2), Root Mean Squared Error (RMSE), and the a_{20} index—were employed to quantify prediction accuracy. Graphs representing the predicted time periods using the models versus the actual time periods are shown in Fig. 14. Table 3 represents a comparison of the evaluation metrics of the aforementioned models with the ML model with the highest prediction i.e., AdaBoost.

To provide a fair baseline comparison that accounts for the additional features available to the ML models, a height-only AdaBoost model was also trained using only building height as the sole input feature. This isolates the incremental predictive value of mass, stiffness, and base width. Table 3 summarises the comparison. The height-only AdaBoost baseline achieves $R^2 = 0.620$ and $RMSE = 0.230$ s, comparable to the empirical code formulas. The full five-feature AdaBoost model achieves $R^2 = 0.969$ and $RMSE = 0.050$ s, demonstrating that the additional structural features (mass, stiffness, base width) provide substantial predictive improvement beyond height alone.

It is also observed from Table 3 that the ML model predicts the time periods with greater accuracy and a lower error compared to the existing models. From the SHAP

analysis conducted, it is observed that the second most influential parameter on the time periods of RC frames is the mass of the frame. None of the empirical models mentioned above accounts for the mass of the frame, and instead calculates the time period as a function of the power of the height.

It is observed from Table 5 that the ML model predicts the time periods with greater accuracy and a lower error compared to the existing models. From the SHAP analysis conducted it is observed that the second most influential parameter on the time periods of RC frames is the mass of the frame. None of the models mentioned above accounts for the mass of the models and instead calculates the time period as a factor of the power of the height.

10 Comparison of Existing Time-Period Estimation with Machine Learning Prediction

To compare the performance of the existing models with the ML models, a fraction (T_{pred}/T_{act}) is calculated. The stochastic distribution of this fraction is depicted via a modified percentile plot in Fig. 16. In this modified representation, the lower boundary of the box corresponds to the 5th percentile, the upper boundary to the 95th percentile, and the central mark to the mean. This representation differs from conventional box-and-whisker plots (which use the 25th and 75th percentile boundaries) and is used here specifically to emphasise tail behaviour and quantify the inherent conservatism embedded within each predictive model, thereby informing its suitability for design-critical applications.

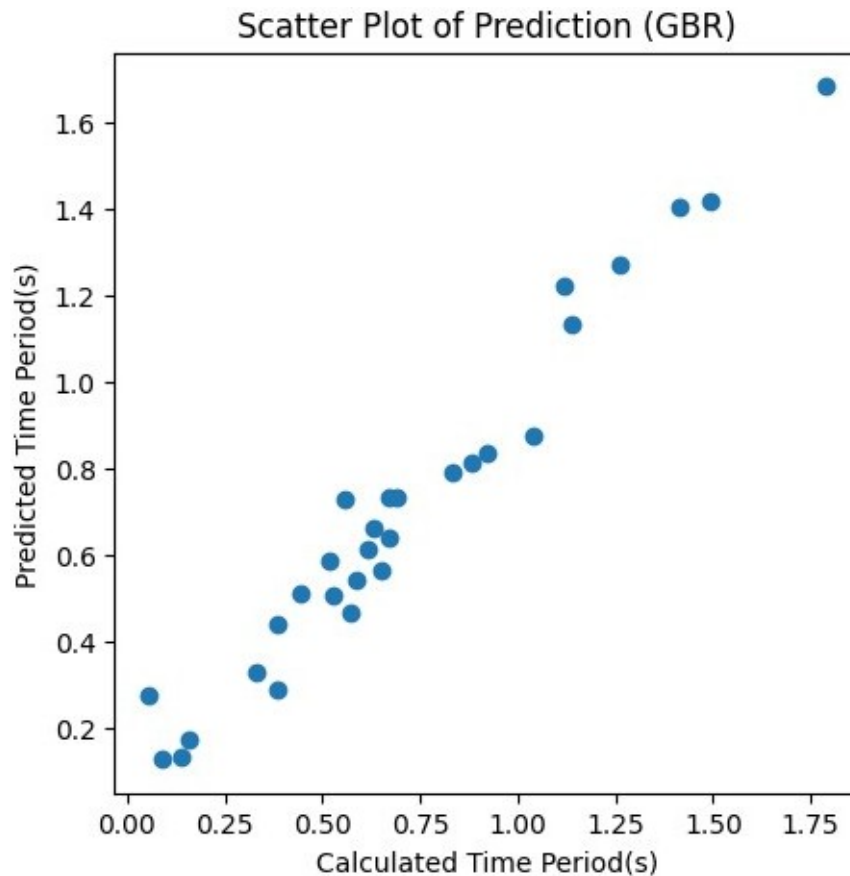


Fig. 10. Predicted by GradBoost vs. Calculated Values of time periods.

Table 5. Comparison of AdaBoost with existing models.

Model	R^2 Score	RMSE	a_{20} index
IS 1893:2016	0.607	0.512	0.212
Height-only AdaBoost (baseline)	0.620	0.230	0.241
Inel et al. (2019)	0.608	0.26	0.237
Goel and Chopra (1997)	0.608	0.41	0.245
Velani and Ramancharla (2023)	0.606	0.26	0.227
AdaBoost	0.969	0.05	0.90

The 5th percentile value indicates that out of 100 predictions, the model should not predict the time-period lesser than the observed value more than five times. This modified representation enables a quantitative assessment of the inherent conservatism embedded within each predictive model, thereby informing its suitability for design-critical applications. The 20th percentile values are also reported in Table 6, as the 5th percentile values may be too conservative. From Table 4, Fig. 15 and 16, it is observed that the mean fraction as per IS 1893:2016 (2016) and Inel et al. (2019) method is 1.97 and 1.12, respectively, indicating that these methods give an overestimate of the actual time period. However, Goel and Chopra (1997) and Velani and Ramancharla (2023) underestimate of the actual time periods. However, the scatter of the fractions in all the theoretical models were extremely high. A similar scatter is also observed in the linear regression model and the support vector regression model. However, all the ML models

have a mean fraction ranging from 1.01 to 1.07, indicating an excellent performance in predicting the time-periods. In addition there is negligible scatter of the predictions. Hence, this method may be used in the evaluation of existing RC frames.

11 Conclusions

This study demonstrates the efficacy of machine learning in predicting the fundamental time periods of regular RC bare-frame buildings. Ensemble methods, particularly AdaBoost, offer superior accuracy and interpretability compared to traditional empirical formulas. SHAP analysis provides valuable insights into feature contributions, enhancing model transparency.

A synthetic dataset of 292 regular RC frame models was generated using the ETABS API, varying parameters like height, base width, mass, and stiffness. Five input features were used: height, base width, height-to-base ratio, mass,

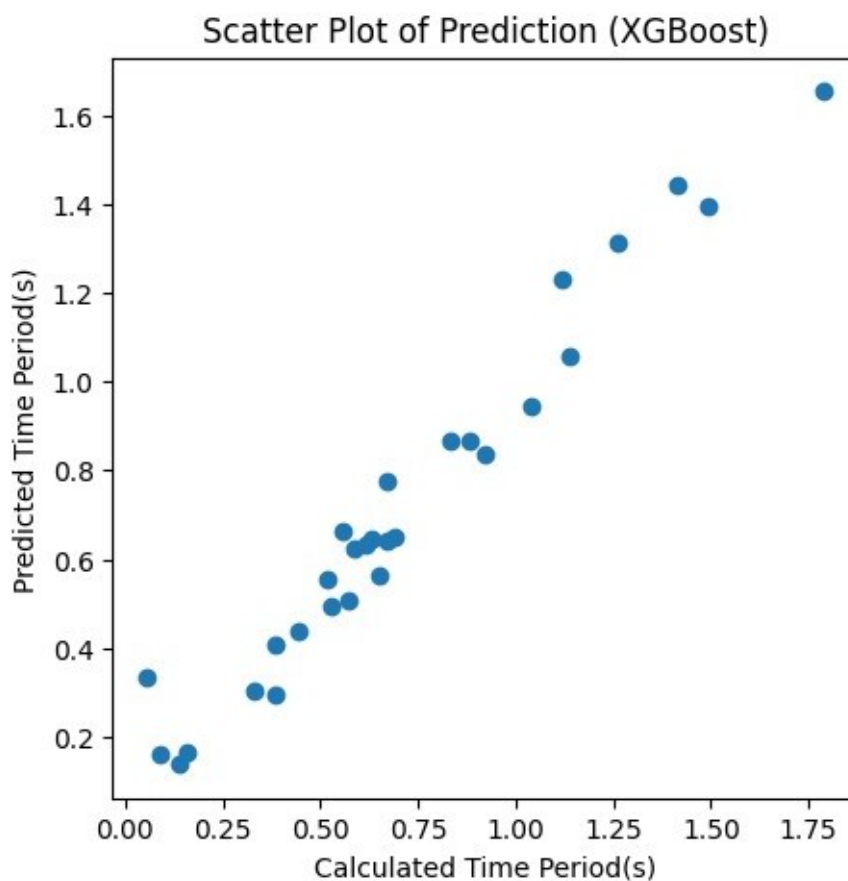


Fig. 11. Predicted by XGBoost vs. Calculated Values of time periods.

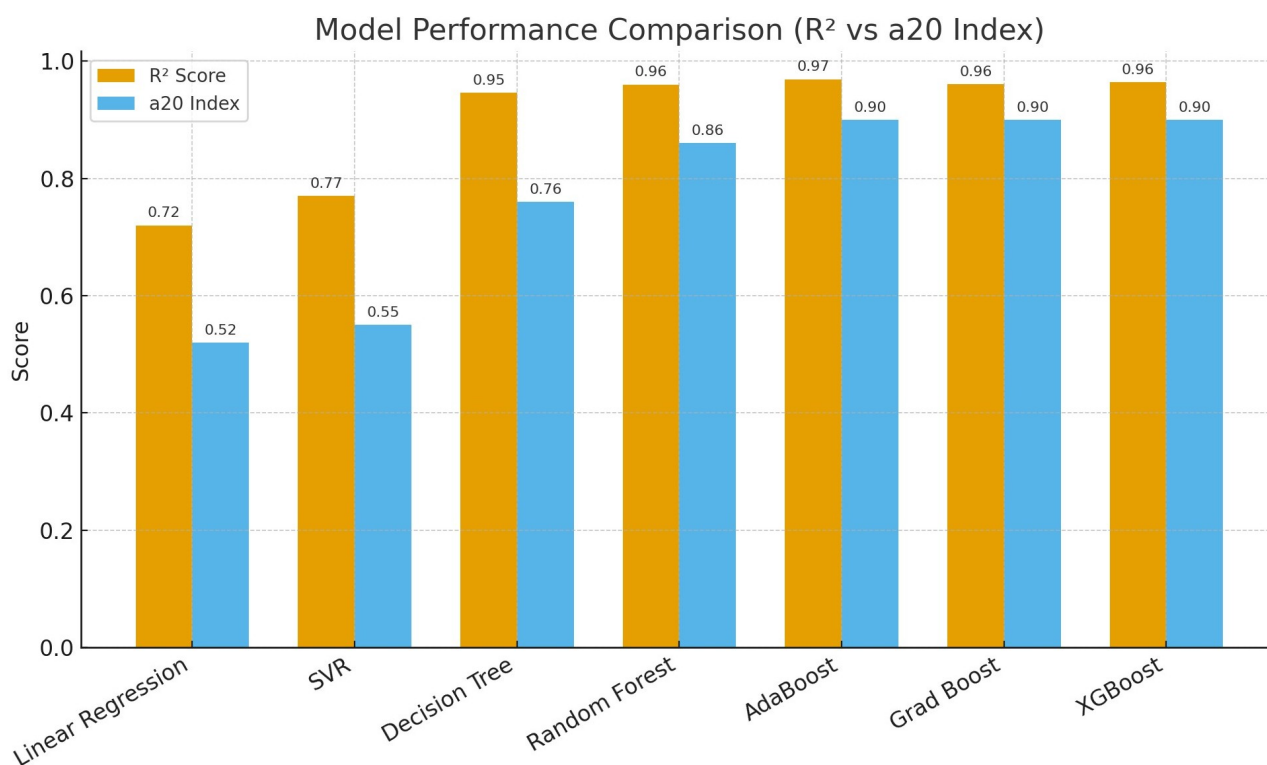


Fig. 12. Comparison of model evaluation using R^2 and a_{20} for the ML models.

and stiffness. The time periods of the RC frames were predicted using seven machine learning algorithms, and the performance of the models was evaluated using R^2 , the a_{20} index, RMSE, and MSE. A review of the ML algorithms

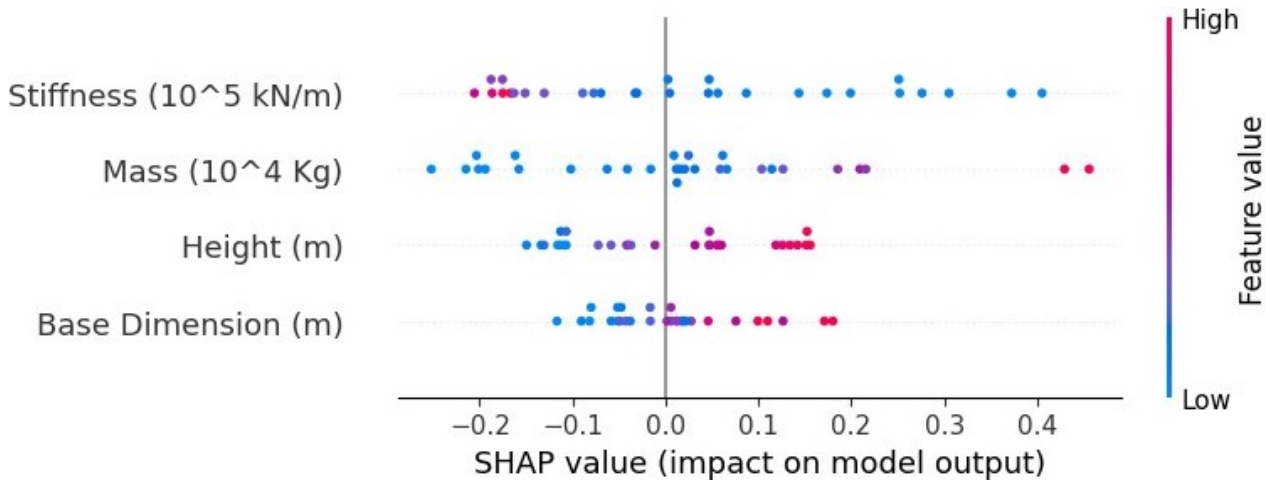


Fig. 13. Shapley value plot for AdaBoost.

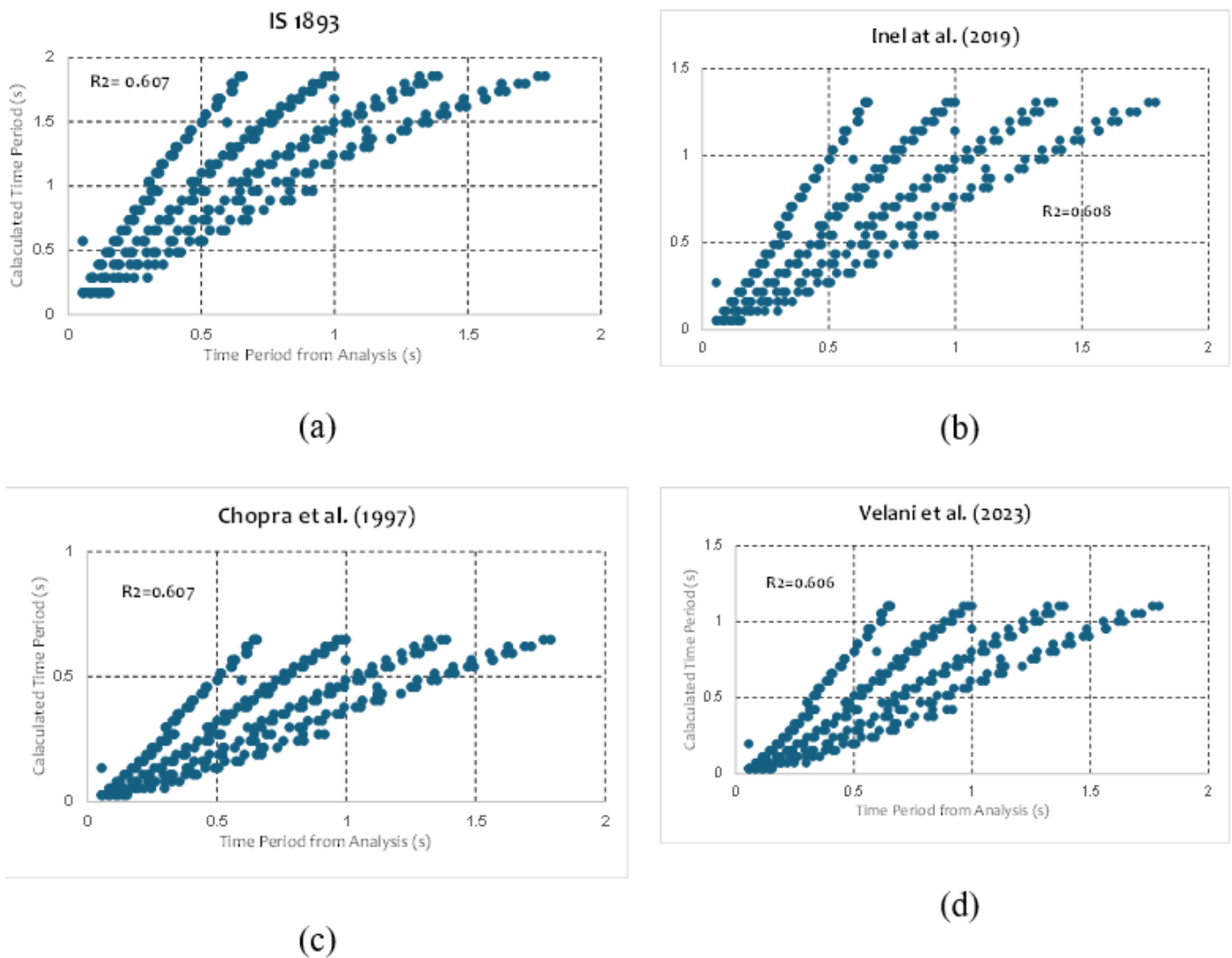


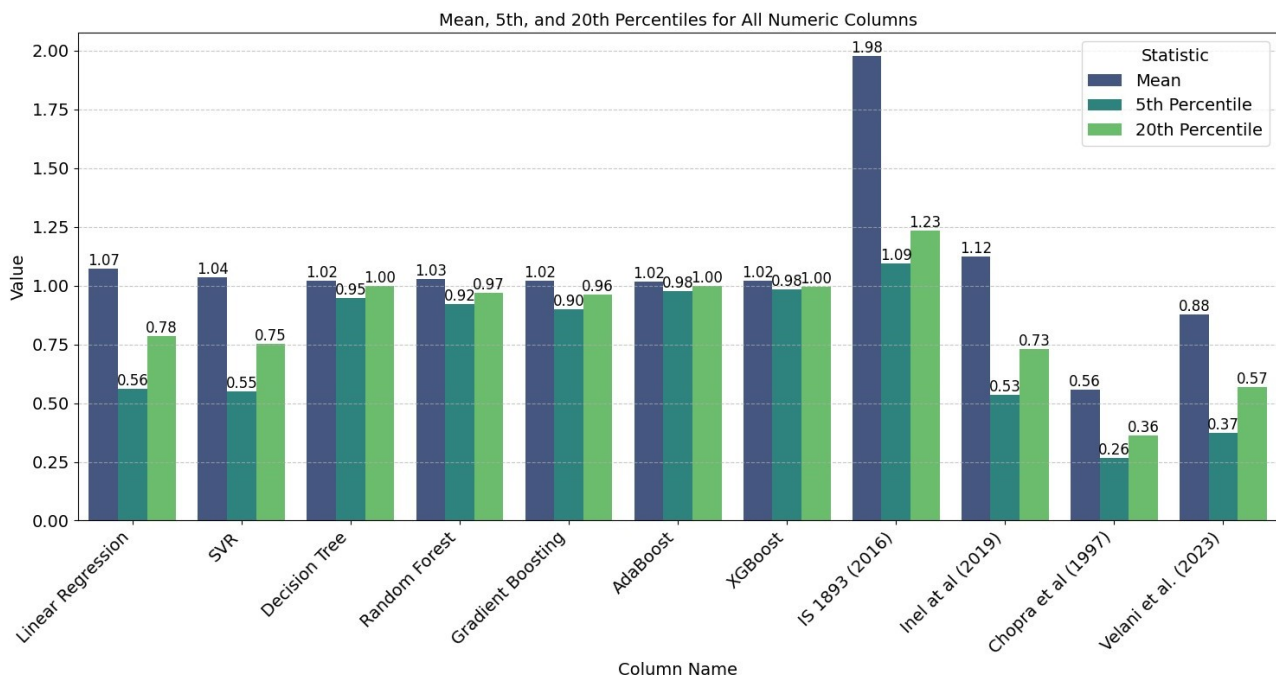
Fig. 14. Scatter plots of time periods predicted by empirical models vs. values calculated by ETABS. (a) IS 1893:2016, (b) Inel et al. (2019), (c) Goel and Chopra (1997), (d) Velani and Ramancharla (2023).

indicates that:

- Linear Regression and SVR offer the lowest complexity but also the poorest fit ($R^2 \approx 0.72$ – 0.77). Errors (MAE ~ 0.14 s) remain relatively high.
- Decision Tree shows a significant jump in accuracy ($R^2 = 0.946$) demonstrating the power of non-parametric splits; however, mild overfitting is evident from the train-test gap.

Table 6. Statistical Analysis of the w_{cr_pred}/w_{cr_act} fraction.

Model	Mean Fraction	5 th percentile	20 th percentile
Linear Regression	1.07	0.56	0.78
SVR	1.03	0.54	0.75
Decision Tree	1.01	0.57	1.0
Random Forest	1.02	0.92	0.96
AdaBoost	1.02	0.89	0.96
GradBoost	1.016	0.97	1.0
XGBoost	1.019	0.98	0.99
IS 1893:2016 (2016)	1.97	1.09	1.23
Inel et al. (2019)	1.12	0.53	0.72
Goel and Chopra (1997)	0.55	0.26	0.36
Velani and Ramancharla (2023)	0.87	0.37	0.58

**Fig. 15.** Statistical analysis of the Time Period Predicted to Time Period Actual fraction.

- Bagging Ensemble (Random Forest): Aggregating trees reduces variance; R^2 improves to 0.962 and MAE drops to 0.061 s. The a_{20} index climbs to 0.86.
- Boosting Ensembles: AdaBoost achieves the best overall fit ($R^2 = 0.969$) and lowest MAE (0.050 s). Gradient Boosting and XGBoost closely follow ($R^2 \sim 0.961$ – 0.964). All three boosting methods share the top a_{20} index (0.90), reflecting exceptional reliability.

Linear and SVR models can serve as quick baseline checks but lack the fidelity required for critical seismic assessments. Boosting models (especially AdaBoost) deliver the most accurate and consistent time-period estimates, making them ideal for integration into seismic design workflows. Random Forest is a strong alternative when training speed or interpretability of individual trees is prioritised over absolute top accuracy. Interpretability-based SHAP analysis enhances engineering trust by providing a clear feature-impact breakdown to validate ML results.

Comparison against existing theoretical models and a height-only ML baseline confirms that the additional structural features (mass, stiffness, base width) are responsible for the superior performance of the multi-feature ML models. Theoretical models tend to either overestimate or underestimate actual time periods with high scatter, while advanced ML models achieve mean prediction ratios close to 1.0 (1.01–1.07) with negligible scatter, indicating high accuracy and reliability for evaluating existing RC frames.

From a disaster risk reduction perspective, the ML models developed in this study provide a practical tool for rapid seismic period estimation of existing regular RC bare-frame buildings, which can support building stock vulnerability assessments and pre-disaster screening programmes. The adoption of such tools can strengthen evidence-based decision-making in seismic hazard management.

The following limitations of this study are acknowledged: (1) The dataset comprises 292 regular bare-frame RC buildings without infill walls or soil-structure interac-

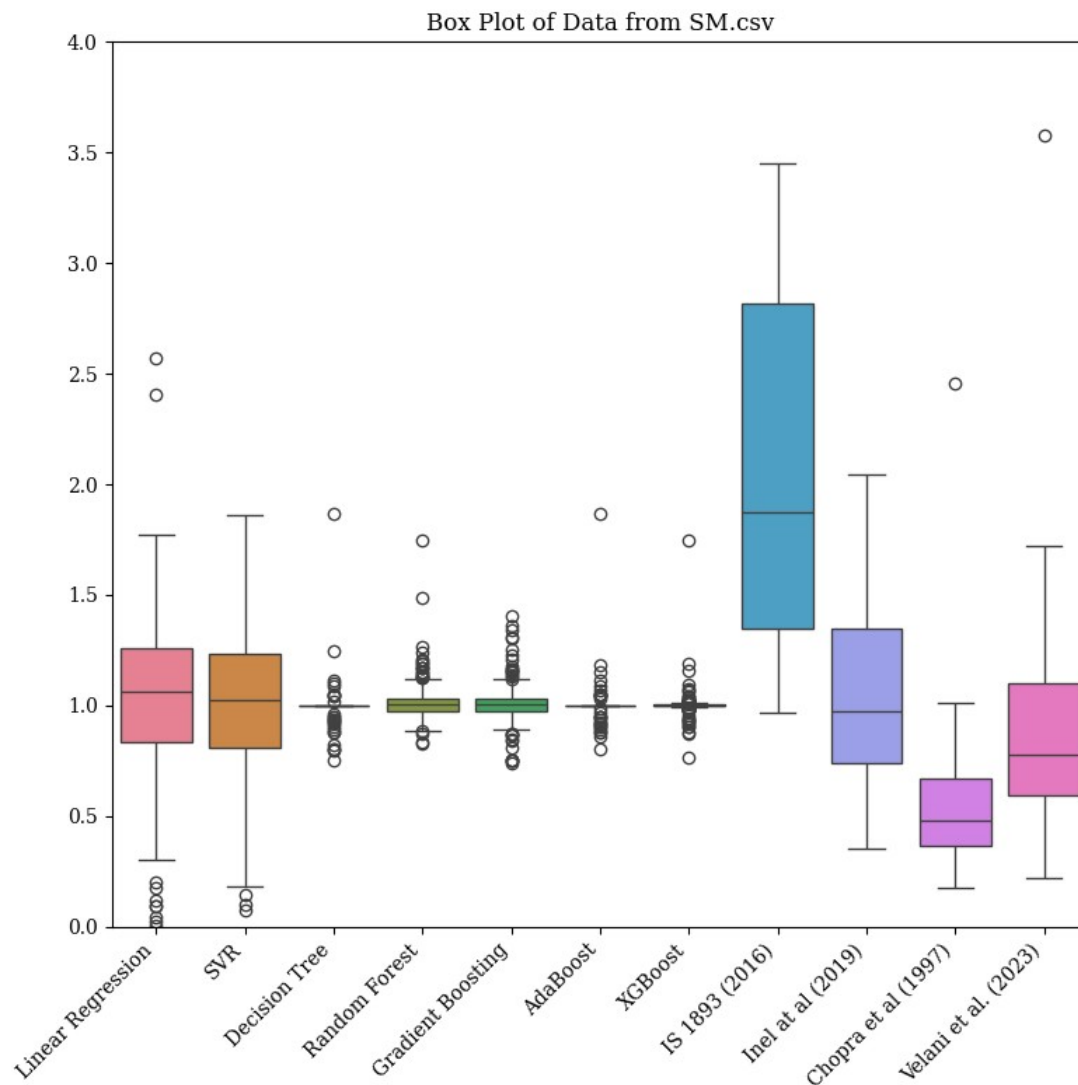


Fig. 16. Modified percentile plot of the $T_{\text{pred}}/T_{\text{act}}$ fraction for all models (box boundaries = 5th and 95th percentiles; central mark = mean).

tion effects; the models are therefore not directly applicable to infilled frames, irregular configurations, or structures on soft soils. (2) Lateral stiffness is used as an input feature but is not available a priori at the design stage, restricting direct application to new building design; the models are best suited for assessment of existing structures. (3) The relatively small dataset (292 samples) limits generalisation to a broader range of structural typologies. Future work should expand the dataset to include infilled frames, irregular configurations, SSI effects, and post-earthquake damage states, and should explore neural network architectures for larger datasets.

The integration of ML models into structural design and assessment workflows can improve seismic safety and design economy. Future work may explore neural network architectures and expand datasets to include more complex structural configurations.

12 Declarations

Data Availability Statement

The dataset generated and used in this study can be obtained by mailing the corresponding author. The dataset includes all 292 RC frame configurations with their structural parameters (height, base width, height-to-base ratio, mass, stiffness) and computed fundamental time periods. It is available without restriction for academic and research purposes.

Ethical approval

The authors confirm that this manuscript or any sections of this is under review anywhere other than this journal. The authors do not require any ethical approval for the dataset published or for the results based on the analysis.

CRedit author statement

Jacob Alex Kollerathu: Conceptualisation, Methodology, Software, Data Curation, Writing – Original Draft, Supervision. Aniket N Bhalkikar: Investigation, Visualisation, Writing – Review and Editing.

AI Usage Statement

The authors confirm that no AI tool has been used as a writing assistant and image generation.

Acknowledgement of Funds

The authors confirm that no funds have been received for this research work.

References

- Ahmad, F., Alasskar, A., Samui, P., Asteris, P., 2025. Machine learning-based graphical user interface for predicting high-performance concrete compressive strength: Comparative analysis of Gradient Boosting Machine, Random Forest, and Deep Neural Network Models. *Frontiers of Structural and Civil Engineering* 19, 1075–1090. <https://doi.org/10.1007/s11709-025-1201-8>.
- Ajay Kumar, S., Pradeep Kumar, R., 2013. Earthquake behavior of reinforced concrete framed buildings on hill slopes, in: *International Symposium on New Technologies for Urban Safety of Mega Cities in Asia (USMCA 2013)*, Hanoi, Vietnam.
- Ali, D., Mahmood, R., Ahmed, M., 2025. Effect of steel bracing on seismic performance of reinforced concrete structures. *SSRG International Journal of Civil Engineering* 12, 96–104. <https://doi.org/10.14445/23488352/IJCE-V12I7P108>.
- Arlekar, J.N., Murty, C.V., 2000. Ambient vibration survey of RC-frame buildings having brick masonry infill walls. *Indian Concrete Journal* 74, 581–588.
- ASCE7-22, 2022. *Minimum Design Loads and Associated Criteria for Buildings and Other Structures*. American Society of Civil Engineers (ASCE), Washington, USA.
- Badkoubeh, A., Massumi, A., 2017. Fundamental period of vibration for seismic design of concrete shear wall buildings. *Scientia Iranica* 24, 1010–1016. <https://doi.org/10.24200/sci.2017.4084>.
- Balendra, T., Ma, Z., Tan, C., 2003. Design of tall residential buildings in Singapore for wind effects. *Wind and Structures* 6, 221–248. <https://doi.org/10.12989/was.2003.6.3.221>.
- Banik, A., Chakma, T., Debbarma, P., Das, B., Sapu, R., Nath, J., Tripura, B., 2025. Effect of floor area ratio reduction on the fundamental period of RC buildings on flat and sloped terrains. *Procedia Structural Integrity* 70, 604–610. <https://doi.org/10.1016/j.prostr.2025.07.096>.
- Bioud, N.-I., Oulad Laid, I., Benbouras, M., 2023. Estimating the fundamental period of infilled RC frames structures via deep learning. *Urbanism. Architecture. Constructions* 14, 59–80.
- Breiman, L., 1996. Bagging predictors. *Machine Learning* 24, 123–140. <https://doi.org/10.1007/BF00058655>.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Wadsworth & Brooks/Cole, Monterey, CA.
- Calitz, D., Markou, G., Bakas, N., Papadrakakis, M., 2023. Developing fundamental period formulae for steel framed structures through machine learning and automated algorithms, in: *COMPDYN Proceedings*. <https://doi.org/10.7712/120123.10468.20613>.
- Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 785–794. <https://doi.org/10.1145/2939672.2939785>.
- Chengsheng, T., Huacheng, L., Bing, X., 2017. AdaBoost typical algorithm and its application research. *MATEC Web of Conferences* 139, 00222. <https://doi.org/10.1051/mateconf/201713900222>.
- Cinar, O.F., Aldemir, A., Zervent, A., Yucel, O.B., Erberik, M.A., Anil, O., Askan, A., 2024. Fundamental period estimation of RC buildings by considering structural and non-structural damage distributions through neural network. *Neural Computing and Applications* 24, 1329–1350. <https://doi.org/10.1007/s00521-023-09107-6>.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine Learning* 20, 273–297. <https://doi.org/10.1007/BF00994018>.
- Di Sarno, D., Amiri, S., 2019. Period elongation of deteriorating structures under mainshock-aftershock sequences. *Engineering Structures*, 196. <https://doi.org/10.1016/j.engstruct.2019.109341>.
- Dietterich, T.G., 2000. Ensemble methods in machine learning, in: *International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science*, 1857, p. 1–15. https://doi.org/10.1007/3-540-45014-9_1.
- Dilmaç, H., 2020. Influence of openings of infill wall on seismic vulnerability of existing RC structures. *Structural Engineering and Mechanics* 75, 211–227. <https://doi.org/10.12989/sem.2020.75.2.211>.
- Ditommaso, R., Vona, M., Gallipoli, M., Mucciarelli, M., 2013. Evaluation and considerations about fundamental periods of damaged reinforced concrete buildings. *Natural Hazards and Earth System Science* 13, 1903–1912. <https://doi.org/10.5194/nhess-13-1903-2013>.
- Elgohary, H.A., 2025. A comprehensive review and ANN analysis of empirical formulae for determining the fundamental period of multi-story RC moment resisting frames. *Journal of Umm Al-Qura University for Engineering and Architecture* <https://doi.org/10.1007/s43995-025-00192-3>.
- Eurocode 8, 2004. *Design of Structures for Earthquake Resistance – Part 1: General Rules, Seismic Actions and Rules for Buildings*. European Committee for Standardization (CEN), Brussels, Belgium.
- Gallipoli, M.R., Petrovic, B., Calamita, G., Tragni, N., Scaini, C., Barnaba, C., Parolai, S., 2023. Towards specific T–H relationships: FRIBAS database for better characterization of RC and URM buildings. *Bulletin of Earthquake Engineering* 21, 2281–2307. <https://doi.org/10.1007/s10518-022-01594-7>.
- Giri, P., Bhatt, A., Gautam, D., Chaulagain, H., 2019. Comparison between the seismic codes of Nepal, India, Japan and EU. *Asian Journal of Civil Engineering* 20, 301–312. <https://doi.org/10.1007/s42107-018-0102-8>.
- Goel, R.K., Chopra, A.K., 1997. Period formulas for moment-resisting frame buildings. *Journal of Structural Engineering* 123, 1454–1461. [https://doi.org/10.1061/\(ASCE\)0733-9445\(1997\)123:11\(1454\)](https://doi.org/10.1061/(ASCE)0733-9445(1997)123:11(1454)).

- Guo, J., Enokida, R., Li, D., Ikago, K., 2023. Combination of physics-based and data-driven modeling for nonlinear structural seismic response prediction through deep residual learning. *Earthquake Engineering & Structural Dynamics* 52, 2295–2571. <https://doi.org/10.1002/eqe.3863>.
- Hadzima-Nyarko, M., Morić, D., Draganić, H., Štefić, T., 2015. Comparison of fundamental periods of reinforced shear wall dominant building models with empirical expressions. *Tehnički Vjesnik* 22, 685–694. <https://doi.org/10.17559/TV-20140228124615>.
- Inel, M., Ozmen, H.B., Cayci, B.T., 2019. Determination of period of RC buildings by the ambient vibration method. *Advances in Civil Engineering*, 1213078. <https://doi.org/10.1155/2019/1213078>.
- IS 1893:2016, 2016. *Criteria for Earthquake Resistant Design of Structures – Part 1: General Provisions and Buildings*. Bureau of Indian Standards (BIS), New Delhi, India.
- Jitth, K., Song, Y., Wang, Z., Li, Y., Chen, J., 2025. A robust automated machine learning predictive model for natural period of buildings. *Structural Engineering and Mechanics* 94, 351–362. <https://doi.org/10.12989/sem.2025.94.5.351>.
- Jose, A., Mishra, N., 2021. Seismic analysis of multistorey RC building with vertical setback and its retrofit strategies. *Lecture Notes in Civil Engineering*, 143. https://doi.org/10.1007/978-981-33-6969-6_42.
- Joseph, H., Quaranta, G., Carboni, B., Lacarbonara, W., 2024. Deep learning architectures for data-driven damage detection in nonlinear dynamic systems under random vibrations. *Nonlinear Dynamics* 112, 20611–20636. <https://doi.org/10.1007/s11071-024-10270-1>.
- Kewate, S., Murudi, M., 2018. A critical study on empirical period–height relationship for reinforced concrete moment resisting frame buildings. *Indian Concrete Journal* 92, 89–94. <https://doi.org/10.1088/1757-899X/481/1/012018>.
- Li, S., Jin, X., Laima, S., Li, H., 2024. Efficient data-driven nonlinear system identification for structural health monitoring: A proof-of-principle study. *Advances in Structural Engineering* 27, 2950–2961. <https://doi.org/10.1177/13694332241281529>.
- Ma, G., Wang, Y., 2023. Prediction of deformation capacity of rectangular RC shear walls subjected to bending damage based on symbolic regression. *Earthquake Engineering and Engineering Dynamics* 43, 139–149. <https://doi.org/10.13197/j.eeed.2023.0615>.
- Masi, A., Vona, M., 2010. Experimental and numerical evaluation of the fundamental period of undamaged and damaged RC framed buildings. *Bulletin of Earthquake Engineering* 8, 643–656. <https://doi.org/10.1007/s10518-009-9136-3>.
- Mohamed, A., El Kashif, K., Salem, H., 2019. An investigation of the fundamental period of vibration for moment resisting concrete frames. *Civil Engineering Journal (Iran)* 5. <https://doi.org/10.28991/cej-2019-03091438>.
- Narkhede, J., 2024. Comparative evaluation of post-hoc explainability methods in AI: LIME, SHAP, and Grad-CAM, in: *4th International Conference on Sustainable Expert Systems, ICSES 2024 Proceedings*, p. 826–830. <https://doi.org/10.1109/ICSES63445.2024.10762963>.
- Nguyen, D., Choo, Y., Nguyen, P., Kim, D., 2025. Nonlinear seismic soil-structure interaction analysis of shallow-founded structure: centrifuge test, numerical analysis, and deep learning approach. *Computers and Geotechnics*, 186. <https://doi.org/10.1016/j.compgeo.2025.107382>.
- Nguyen, H., Kim, C., Lee, K., Shin, M., 2024. Development of data-driven models to predict seismic drift response of RC wall structures: An application of deep neural networks. *Soil Dynamics and Earthquake Engineering*, 186. <https://doi.org/10.1016/j.soildyn.2024.108952>.
- Panzer, F., Lombardo, G., Muzzetta, I., 2013. Evaluation of building dynamic properties through in situ experimental techniques and 1D modeling: The example of Catania, Italy. *Physics and Chemistry of the Earth, Parts A/B/C* 63, 136–146. <https://doi.org/10.1016/j.pce.2013.04.008>.
- Rahman, T., Momin, M., Provash, A., 2025. Comprehensive analysis of structural parameters influencing the fundamental period of steel-braced RC buildings using machine learning interpretability. *AI in Civil Engineering* 4. <https://doi.org/10.1007/s43503-025-00051-z>.
- Ron, K., 1995. A study of cross validation and bootstrap for accuracy estimation and model selection, in: *International Joint Conference on Artificial Intelligence (IJCAI)*, Montreal, Quebec.
- Sarma, S., Sundar Das, T., Bora, A., Bharadwaj, K., 2022. Parametric study on the variation of time period of RC MRF buildings. *Lecture Notes in Civil Engineering* 175, 415–426. https://doi.org/10.1007/978-981-16-4617-1_33.
- Schapire, R., 2013. Explaining AdaBoost, in: Schölkopf, B., Luo, Z., Vovk, V. (Eds.), *Empirical Inference*. Springer, p. 37–52. https://doi.org/10.1007/978-3-642-41136-6_5.
- Shan, J., Huang, C., Wang, L., Loong, C., 2024. Data-driven prediction of natural period for existing RC high-rise buildings using probabilistic machine learning methods. *Journal of Building Engineering*, 90. <https://doi.org/10.1016/j.jobbe.2024.109394>.
- Sharma, N., Dasgupta, K., Dey, A., 2023. Soil-structure interaction induced modification on the natural period of reinforced concrete buildings. *Lecture Notes in Civil Engineering* 326, 352–361. https://doi.org/10.1007/978-3-031-30125-4_32.
- Shendkar, M., Beiraghi, H., Mandal, S., 2021. Effect of irregularity on seismic design parameters of RC-infilled structures. *Magazine of Civil Engineering* 108. <https://doi.org/10.34910/MCE.108.7>.
- Ukwaththa, U., Liyanarachchi, T., Kulasoorya, W., Ranasinghe, R., 2024. Use of interpretable machine learning methods to predict the fundamental period of masonry infilled reinforced concrete frame structures. *Moratuwa Engineering Research Conference (MERCon)*, 127–132. <https://doi.org/10.1109/MERCon63886.2024.10688600>.
- Velani, P., Ramacharla, P.K., 2023. Approximate natural period expression for reinforced concrete tall buildings in India. *Current Science* 124, 1412–1421. <https://doi.org/10.18520/cs/v124/i12/1412-1421>.
- Xiong, W., Jiang, L.-Z., Li, Y.-Z., 2016. Influence of soil-structure interaction (structure-to-soil relative stiffness and mass ratio) on the fundamental period of buildings: Experimental observation and analytical verification. *Bulletin of Earthquake Engineering* 14, 139–160. <https://doi.org/10.1007/s10518-015-9814-2>.
- Zhang, H., Cheng, X., Li, Y., He, D., Du, X., 2023. Rapid seismic damage state assessment of RC frames using machine learning methods. *Journal of Building Engineering*, 65. <https://doi.org/10.1016/j.jobbe.2022.105797>.