




Article

Machine Learning Algorithm for Shear Strength Prediction of Short Links for Steel Buildings

Ghassan Almasabha ^{1,*}, Odey Alshboul ¹, Ali Shehadeh ² and Ali Saeed Almuflih ³

¹ Department of Civil Engineering, Faculty of Engineering, The Hashemite University, Zarqa 13133, Jordan; odey.shboul@hu.edu.jo

² Department of Civil Engineering, Hijjawi Faculty for Engineering Technology, Yarmouk University, Shafiq Irshidatst, Irbid 21163, Jordan; ali.shehadeh@yu.edu.jo

³ Department of Industrial Engineering, King Khalid University, King Fahad St., Guraiger, Abha 62529, Saudi Arabia; asalmuflih@kku.edu.sa

* Correspondence: ghassans@hu.edu.jo

Abstract: The rapid growth of using the short links in steel buildings due to their high shear strength and rotational capacity attracts the attention of structural engineers to investigate the performance of short links. However, insignificant attention has been oriented to efficiently developing a comprehensive model to forecast the shear strength of short links, which is expected to enhance the steel structures' constructability. As machine learning algorithms were successfully used in various fields of structural engineering, the current study fills the gap in estimating the shear strength of short links using sophisticated machine learning algorithms. The deriving factors such as web and flange slenderness ratios, the flange-to-web area ratio, the forces in web and flange, and the link length ratio were investigated in this study, which is imperative to formulate an integrated prediction model. Consequently, the aim of this study utilizes advanced machine learning (ML) models (i.e., Extreme Gradient Boosting (*XGBOOST*), Light Gradient Boosting Machine (*LightGBM*), and Artificial Neural Network (*ANN*)) to produce accurate forecasting for the shear strength. In this study, publicly available datasets were used for the training, testing, and validation. Different evaluation metrics were employed to evaluate the prediction's performance of the used models, such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). The prediction result displays that the *XGBOOST* and *LightGBM* provided better, and more reliable results compared to *ANN* and the AISC code. The *XGBOOST* and *LightGBM* models yielded higher values of R^2 , lower (RMSE), (MAE), and (MAPE) values and have shown to perform more accurately. Therefore, the overall outcomes showed that the *LightGBM* outperformed the *XGBOOST* model. Moreover, the overstrength ratio predicted by the *LightGBM* showed an excellent performance compared to the Gene Expression and Finite Element-based models. The developed models are vital for practitioners to predict the shear strength accurately, which paves the road towards wider application for automation in the steel buildings.

Keywords: shear strength; short link; steel construction industry; machine learning models



Citation: Almasabha, G.; Alshboul, O.; Shehadeh, A.; Almuflih, A.S. Machine Learning Algorithm for Shear Strength Prediction of Short Links for Steel Buildings. *Buildings* **2022**, *12*, 775. <https://doi.org/10.3390/buildings12060775>

Academic Editor: Bo Yang

Received: 9 May 2022

Accepted: 2 June 2022

Published: 6 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Short links are *W* – *shape* steel sections that are either constructed or rolled with link length ratio, $e/(M/V)$, less than 1.6 (AISC, 2016) [1]; where e represents the link length, M and V represent the plastic moments, and shear capacity, respectively. Short links are widely employed in steel bridges, Eccentric Braced Frames (*EBFs*) and coupled walls. The short links have several advantages, such as exceptional plastic rotational capacity and plastic shear capacity [2]. However, several studies observed that the AISC formula underestimates the predicted short links' shear strength [3–7]. The AISC, 2016 [2] Equation (F3–2) estimates the plastic short links' shear strength using $V_p = 0.6 \times F_y (d - 2tf) \times tw$; where V_p represents the plastic shear strength (N), F_y represents the measured steel yield

strength of the web (MPa), d is the link depth (mm), t_f and t_w are the flange and web thicknesses (mm), respectively. Several investigations revealed the major factors that control the shear link strength, such as flange contribution [3,5], cyclic hardening [3], web slenderness [4], and link length ratio [4,6,7].

The testing program of McDaniel et al., 2003 [5], which included two full-scale built-up short links, revealed that in terms of degrading the shear strength of tested links, the cutting-edge factor is a brittle fracture on the linked web. As a result, the tested short links exhibited overstrength factors of 1.83 and 1.94. In addition, Dusicka et al., 2010 [8] investigated the effect of steel yield stress 100, 225, integrated steel strength of 100 and 440, 345, 485 MPa for five plate steel shear links. The obtained results illustrate that the steel links with a low grade attained a plastic rotation of 0.2 rad while the conventional links reached 0.12 rad, and the low-grade steel links achieved an overstrength factor considerably higher than conventional links. Moreover, the average overstrength factor for the tested 12 short links (length ratio ranges from 0.58 to 0.97) was 1.9, Ji et al., 2015 [3]. Furthermore, the very short links reached a plastic rotation of 0.14 rad greater than the 8% limit of AISC 341-10 [9,10]. In addition, Ji et al., 2016 [11] found that the average shear strength of four built-up short links reached 2.0.

Similarly, Liu et al., 2017 [4] noticed that the short links' shear strength was significantly impacted by the web slenderness and link length ratio, and the overstrength ratio for the 12 built-up short links was between 1.35 to 1.5. The link length was critical in the steel links, Okazaki, T. 2004 [7]. The experimental program included 16 link-to-column connections with different link length ratios, where the overstrength ratio varied between 1.05 and 1.47. Bozkurt and Topkaya 2017 [12] discovered that plastic rotation and the overstrength ratio are negatively associated with the link length ratio. The test program included seven short links considering several features (i.e., loading protocol, the link length ratio and stiffeners spacing). In addition, an overstrength ratio of 1.87 to 2.3 was achieved in Bozkurt et al., 2019 [13], where six specimens with a link length of 600 to 800 mm were tested.

To explore the link's ultimate rotational capacity, shear capacity, buckling of flanges, and web, the analysis of complex finite element of shear links was implemented [3,14–20]. However, finite element simulation is considered time consuming, especially in the modeling process and validation of the performance of the predicting model. Moreover, FEA requires special experts in the mechanics of materials and computer aided-software engineering. Recently, the huge availability of databases in the wide range of engineering applications paved the way to extensively and successfully use the machine learning tools to help engineers save the cost, time, and efforts. A leaping use of machine learning has been witnessed in the various civil engineering fields over the last decade. While machine learning has been successfully utilized in the civil engineering applications [21–33], limited studies used machine learning tools to address the shear strength of short links; where a few studies have employed actual experimental databases to validate ML algorithms. Moreover, the dataset applied for forecasting experiments has restricted factors, such as mechanical properties and geometrical dimensions of steel links. The prediction problem for the short links' shear strength was mostly investigated through conventional ML algorithms with limited variables such as FEM and multiple linear regression. To address the mentioned research gaps, the current study proposed to develop the prediction model of the short links' shear strength using innovative systematic and understandable ML algorithms (i.e., XGBOOST, LightGBM, ANN). Moreover, data preprocessing has generated a more precise forecasting model. Furthermore, statistical performance measurements such as (RMSE), (MAE), (MAPE), and (R^2) were employed for comparative analysis. The prediction of the short links' shear strength was formulated in terms of different influence features such as (b_f/t_f), (d/t_w), (A_f/A_w), ($A_f f_{yflange}$), ($A_w f_{yweb}$), and $e/(M/V)$) to overcome the knowledge gap of the available models. Additionally, a comprehensive variables importance analysis has been conducted to investigate the relative impact of these variables on the proposed model.

2. Literature Review

In the related literature on forecasting models, the short links' shear strength was categorized into analytical and ML-based models.

2.1. Analytical Models

The literature includes three analytical models to assess the shear strength of shear links (i.e., AISC 2016 [1], Corte et al., 2013 [15], and G. Almasabha 2022 [34]). The following discussion summarizes the available models.

2.1.1. AISC 2016

The AISC 2016 [1] adopted Equation (1) for the assessment of shear strength of links. It is worth mentioning that Equation (1) does not take into consideration the role of link length proportion, the contribution of flanges, and the slenderness ratio of web or flanges.

$$V_p = 0.6 \times F_y \times (d - 2t_f) \times t_w \quad (1)$$

2.1.2. Corte et al., 2013

A finite element-based algorithm [15] has been proposed to estimate the overstrength ratio ($V_{0.08}/V_y$) of wide flange shear links without axial restraint, where $A_v = (d - t_f)t_w$ and $V_y = (F_y/\sqrt{3})(d - t_f)t_w$. It is worth mentioning that the authors derived Equation (2) for the hot rolled steel link. However, the experimental database of the current study includes both hot rolled and built-up steel links.

$$\frac{V_{0.08}}{V_y} = 1 + 1.35 \left(\frac{A_f}{A_v} \right) \left(\frac{d}{e} \right) \quad (2)$$

2.1.3. G. Almasabha 2022

This study used the gene expression model to build a mathematical equation for the shear link strength (V_{GEP}) [34]. Various parameters were considered in this equation, such as b_f/t_f , d/t_w , A_f/A_w , $A_f f_{yflange}$, $A_w f_{yweb}$, and $e/(M/V)$.

$$V_{GEP} = \left(\frac{1.047}{\left(\frac{e}{M/V} \right)^{0.416}} \right) \times (A_w F_{yweb})^{-0.017} \times \left(\frac{A_f}{A_w} \right)^{0.12} \times \left(A_w F_{yweb} - \left(\frac{d}{t_w} \right)^{0.6} + \left(\frac{e}{M/V} \right)^{0.2} + \frac{b_f}{t_f} \right) \quad (3)$$

2.2. ML Models

Although innovative ML algorithms outperform traditional models in most research disciplines [35,36], ML models are still humbly utilized to forecast the short links' shear strength such as [15–18,37–39], also limited variables with humble datasets have been applied in these studies. On the other hand, these studies' executed ML models seem somewhat traditional. Therefore, the need to use sophisticated ML models for shear strength prediction of short links to reduce prediction error with better accuracy is becoming required. The XGBOOST technique was recently developed using a tree-based ensemble, a complex gradient boosting with higher processing abilities, and an excellent tool to deal with over-fitting concerns [40]. The learning method uses a boosting framework-based decision tree called *LightGBM*, released by Microsoft in 2017. It is quicker, uses less memory, and is more accurate than XGBOOST [41]. *LightGBM* also provided decision rules for category features, which transform factors into one-time multidimensional functionality, saving time and memory [42].

3. Methodology

The aim of this research is to promote an algorithm for predicting the short links' shear strength accurately. In addition, various statistical testing (e.g., data cleaning, normalization, and standardization) were also employed for the purpose of assessing the model's validity and reasonableness. Moreover, data collection and feature definition, data preprocessing, ML algorithm, and model performance evaluation are the four primary aspects of the suggested methodology. The methodology flowchart is shown in Figure 1.

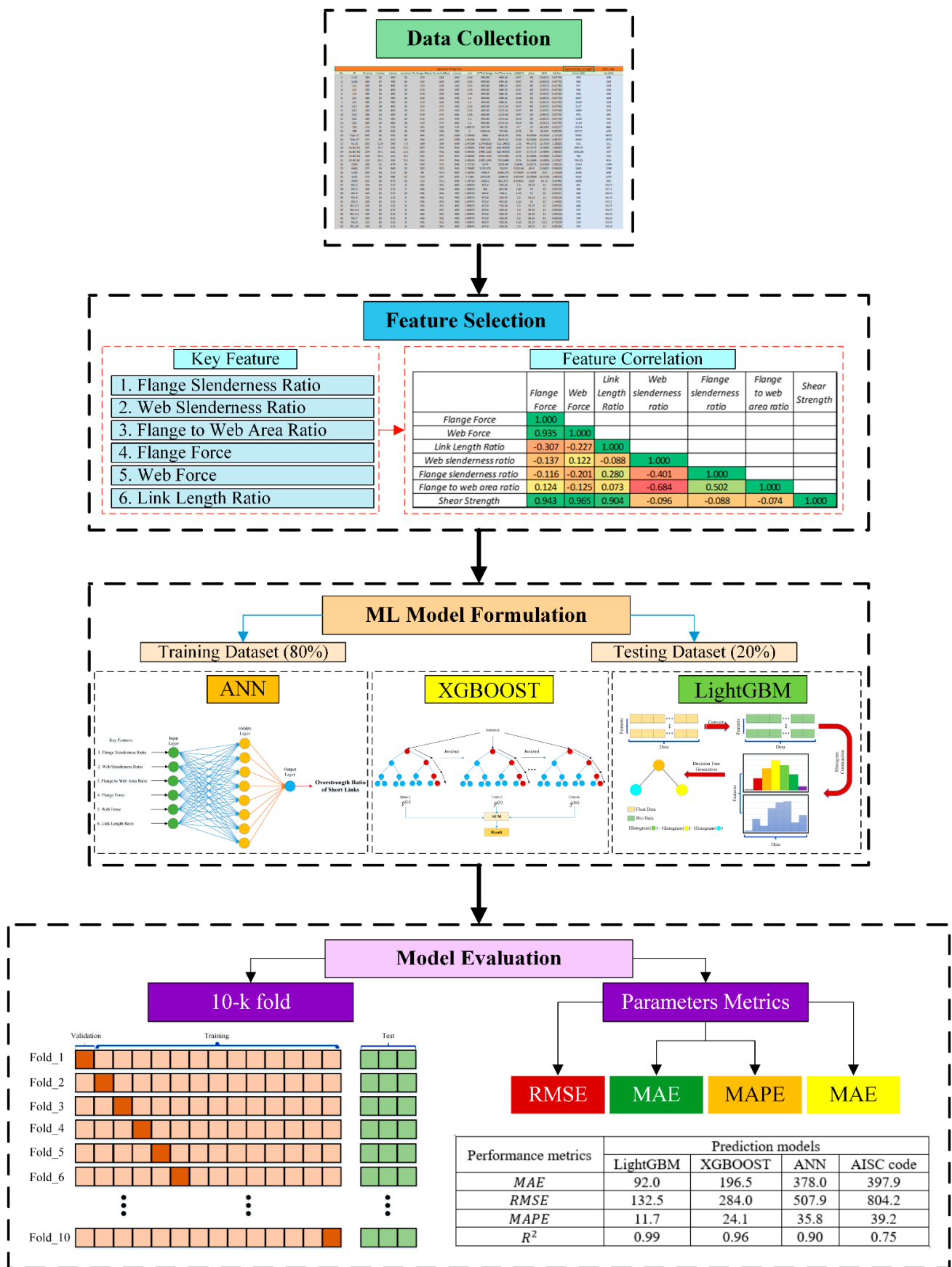


Figure 1. Processing structure for prediction of the short links' shear strength.

3.1. Data collection and Feature Definition

The experimental data of 110 samples were gathered from 1983 to 2019 [3,5–8,11–13,43–46] to evaluate the features impacting the short links' shear strength under natural settings. Several significant features have been considered in this study, such as (b_f/t_f) , (d/t_w) , (A_f/A_w) , $(A_f f_{yflange})$, $(A_w f_{yweb})$, and $e/(M/V)$. Table 1 illustrates the model variables along with their description.

Table 1. Description of model variables.

Feature	Definition	Data Type
(b_f/t_f)	Flange slenderness ratio	Numeric
(d/t_w)	Web slenderness ratio	Numeric
(A_f/A_w)	Flange to web area ratio	Numeric
$(A_f f_{yflange})$	Flange force	Numeric
$(A_w f_{yweb})$	Web force	Numeric
$e/(M/V)$	Link length ratio	Numeric

3.2. Data Preprocessing

Data preprocessing is a key factor in managing the dataset before employing ML algorithms. The data preparation procedure also increases the model's prediction performance. Data noise, removing outliers, normalization, and standardization are among the approaches' processes. All the related variables had numerical values, as shown in Table 1. The initial stage in data preparation is to eliminate outliers from the collected data set. There would be uncertainty and inaccuracy in the data set because of the outlier, which might reduce the efficacy of the linear regression technique. Interquartile ranges were performed to analyze extreme and outlier results in this research. Outliers were removed using graphical techniques such as Boxplots. To characterize and eliminate residuals from the acquired data, null indicators have been employed. Reliability discomfiture might happen when some data are lost from the initial database. It could be concluded that the missing data (values represented by the "Null" or "-" indicators) were taken into account. Preprocessing, which includes eliminating outliers and normalizing the dataset, is required once the necessary variables have been identified. Data standardization was implemented in the current study, where no outliers in the dataset need to be removed.

3.3. ML Algorithm

To forecast the short links' shear strength, three ML algorithms have been utilized (i.e., ANN, XGBOOST, and LightGBM). Training and testing sets are extracted from the database to ensure the proposed ML algorithm's effectiveness. The data was divided into training and testing with a proportion of 80% and 20%, respectively. The suggested prediction models were tested for robustness and effectiveness using 10-fold cross-validation. The model characteristics and implementation procedure are depicted in the next section.

3.3.1. Artificial Neural Network

ANN is a neural network based on an organism's nervous system model. The work principle of a neural network is to create neurons that save, handle data, and connect them with artificial synapses [47]. A plethora of layers of nodes are composed together and communicate with one another to build up a neural network, where the ANN consists of, mainly, an input layer, a hidden layer, and output layers. There is a direct relation between the number of nodes in the input layer and the number of variables that illustrate the assessed qualities; on the other hand, the number of neurons in the output layer corresponds to the number of classes. The process is in sequence where the output of the first layer is taken as input to the next layer. The output layer's nodes may provide the desired result when this technique is repeated. The nature of the task and the size of the

data assign the number of neurons and hidden layers. In the hidden and output layers, each neuron is connected to all nodes in the previous layer via a numerical weight. The ANN model has produced the best results, with a max depth of four and learning rate of 0.08. The Neural structure is built up of seven layers: the initial layer is called the input layer and contains the inputs parameters of an ANN with six neurons, the final layer is termed the output layer and contains a single neuron, while the five hidden layers contain eight neurons each. The input, hidden, and output layers of the ANN are shown in Figure 2. The ANN mathematical formulation is provided in the equations below [48].

$$n_k^h = \sum_{j=1}^X w_{kj}^h t_j + b_k^h, \quad k = 1 \text{ to } z \quad (4)$$

where X represents the number of associated features, z denotes the number of hidden neurons, t denotes each input feature, b denotes the hidden layer's bias, and w denotes the weight. The activation function takes the weight of each computed component as an input. The total of the weighted values determines the output.

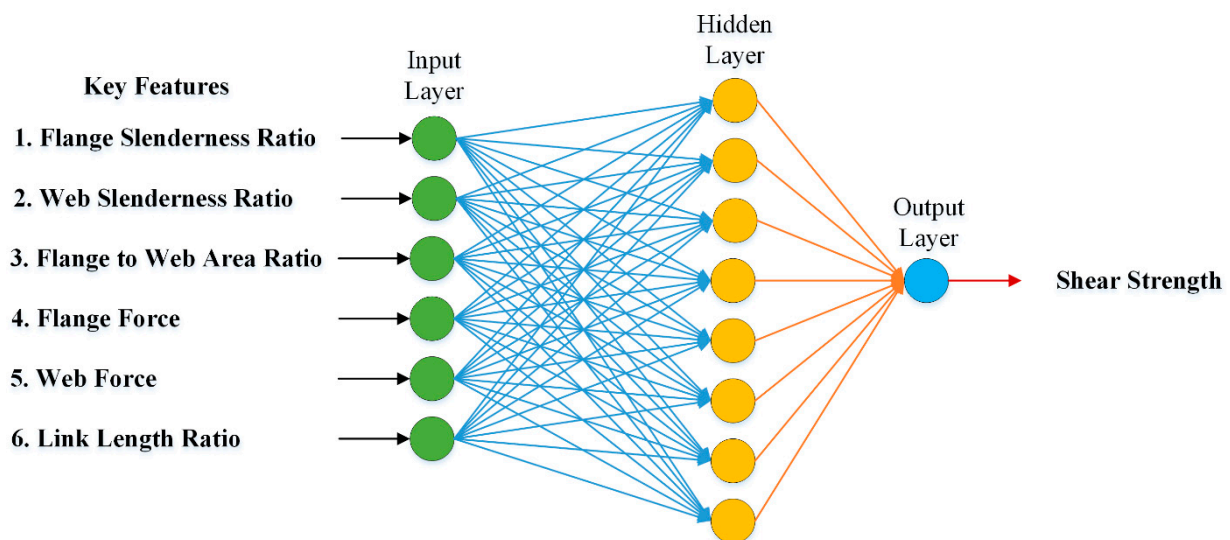


Figure 2. The architecture of the ANN to estimate the short links' shear strength.

3.3.2. Extreme Gradient Boosting

XGBOOST is a newly introduced ML algorithm that has widespread use in several disciplines. This technique is set to be unique in gradient boosting machine and chiefly for regression and classification trees [49]. The XGBOOST is created based on the “boosting” concept, which combines the forecast of weak learners with additive training methodologies to build a strong learner. It aids in the avoidance of over-fitting and enhances computing ability. Figure 3 illustrates the XGBOOST schematic, which simplifies the goal functions, allowing the prediction and regularization terms to be combined while maintaining the fastest feasible processing performance. With a number of trees of 1200, a learning rate of 0.05, and a maximum depth of 15 in XGBOOST, the best results were obtained. The following equations at step p used to determine the common function of the forecast [40].

$$f_i^{(p)} = \sum_{k=1}^p f_k(x_i) = f_i^{(p-1)} + f_p(x_i) \quad (5)$$

where $f_p(x_i)$ represents the learner at step p , $f_i^{(p)}$ represents the prediction at p , $f_i^{(p-1)}$ represents the prediction at $p - 1$, and x_i represents the input features.

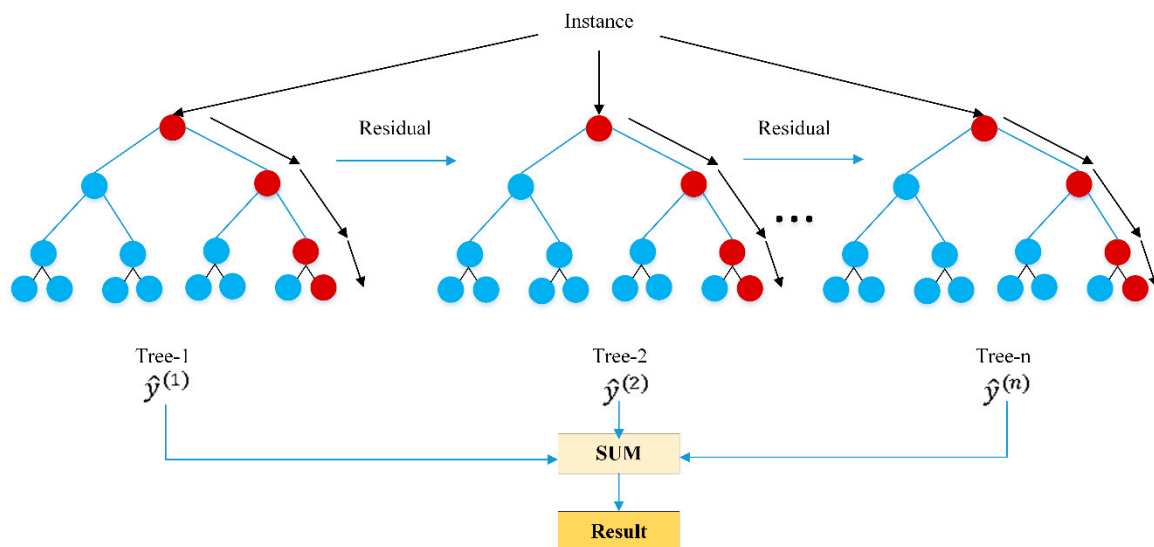


Figure 3. XGBOOST mechanism.

To overcome the issue of over-fitting concerns on the cost of the model's computing speed, the XGBOOST produces analytical formula shown below to assess the "goodness" of the model from the authentic function.

$$Objective^{(p)} = \sum_{k=1}^n l(\bar{y}_i, y_i) + \sum_{k=1}^p \sigma(f_i) \quad (6)$$

where l represents the loss function, n represents the number of data utilized, and σ is the regularization term, which is described in equation below:

$$\sigma(f) = \gamma T + 0.5\lambda \|\omega\|^2 \quad (7)$$

where ω denotes vector scores in leaves, γ denotes the minimal loss necessary to divide the leaf node further, and λ denotes the regularization parameters.

3.3.3. Light Gradient Boosting Machine (LightGBM)

Microsoft Research introduced *LightGBM*, a decision tree with gradient-boosting based on a decision-tree method [50,51]. *LightGBM* is a powerful technique for resolving regression and classification issues. It utilizes low computational memory with a higher accuracy of prediction in comparable to XGBOOST. According to the histogram method and tree leaf-wise growth approach, *LightGBM* enhances the training process and minimizes memory consumption. Figure 4 depicts the histogram decision tree-based algorithm. Figure 4 also depicts the level-wise and leaf-wise development techniques. According to the level-wise growth approach, the leaves on the same layer are divided simultaneously. To control the complexity of the model, it is desirable to optimize using several threads. Furthermore, leaves on the same layer are handled uniformly, despite absorbing different amounts of information. *LightGBM* produced the best results with a tree count of 1100, a learning rate of 0.06, a needed leaf count of 18, and a maximum depth of 13. The information gain (IG) depicts the predicted reduction in entropy caused by dividing nodes into qualities that can be computed in the equations below [40].

$$IG(C, V) = Fn(C) - \sum_{V \in Values(v)} \frac{|C_v|}{C} Fn(C_v)Fn(C) = \sum_{b=1}^B -pb \log_2 pb \quad (8)$$

where $Fn(C)$ represents the input entropy of the group C , pb represents the proportion of C associated with part b , B represents the number of parts, v represents the features

value V , and C_v represents the C 's subset for the features having value v . The process might be affected by several factors, resulting in it being an insignificant process, such as specific leaves with reasonably minimal information gain are discarded, gaining additional memory storage capacity.

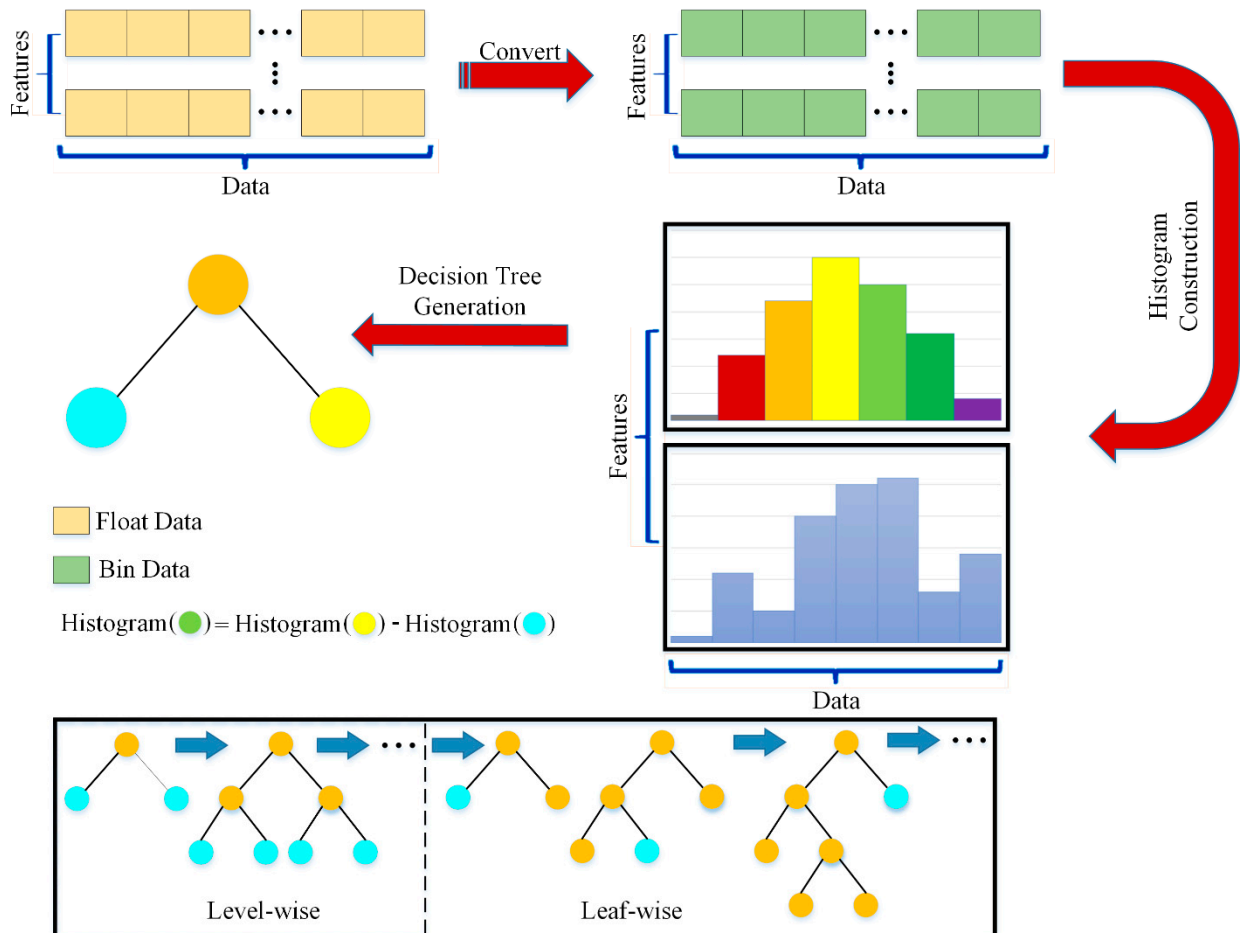


Figure 4. Illustration of *LightGBM* algorithm.

The leaf-wise technique has a better chance of succeeding since it splits the leaf with the largest information gain on a similar layer, as shown in Figure 4. Furthermore, trees with a high depth can be generated using this approach, resulting in an introduction to the maximum depth restriction through tree growth [52].

3.4. Stratified K-Fold Cross-Validation

The K-fold cross-validation approach can enhance the model's accuracy by evaluating how the ML algorithm performs on a new data set. After the datasets have been divided into training and testing subsets, the modeling process is employed for the ML prediction algorithm. The training dataset is divided into several 'k' smaller portions throughout this process. As a result, the moniker 'k'-fold was coined. K-fold is used for testing, and k-1 is used for training based on a random data set. The efficiency of the ML model is examined using a stratified 10-fold cross-validation approach. The data set is divided into ten folds at random using this method. As a result, each of the folds is used just once as a validation set. Finally, the error or accuracy measure for each fold may be compared, and if they are similar, the diversity of the model is high. The 10-fold cross-validation technique is depicted in Figure 5.

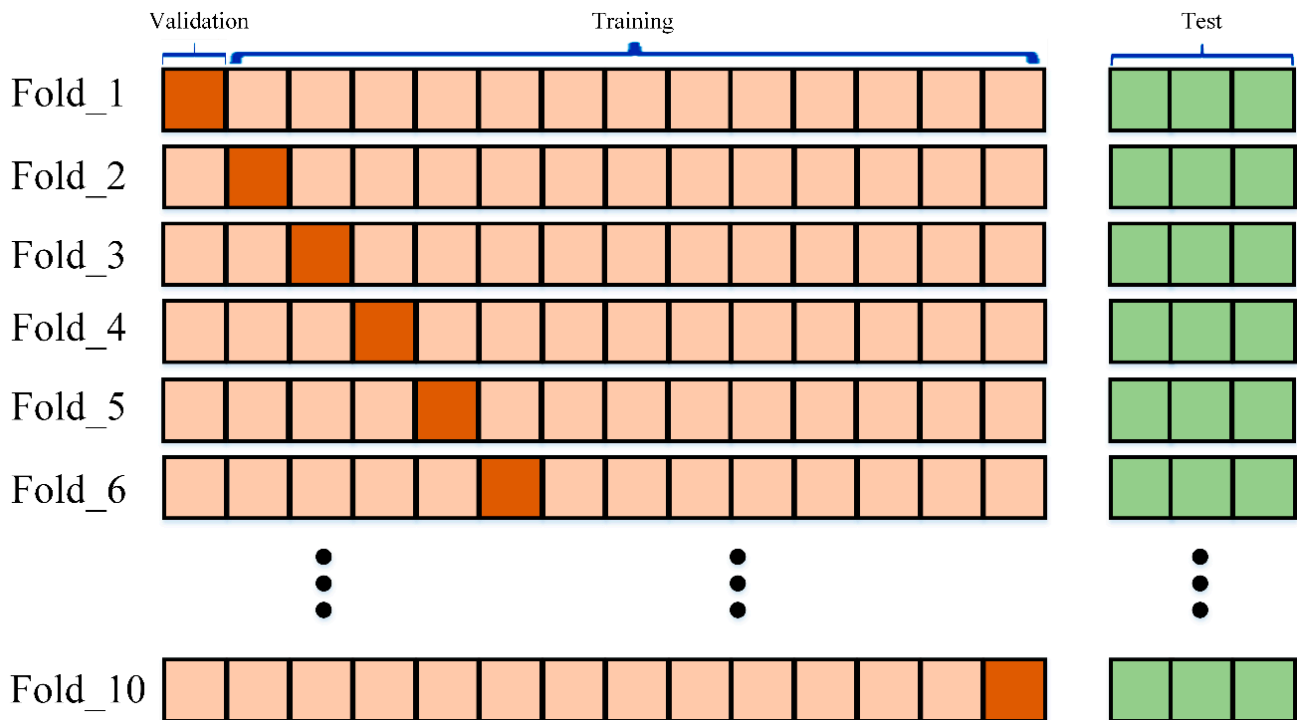


Figure 5. The 10-fold cross-validation technique exemplification.

3.5. Prediction Accuracy Measurement

The evaluation measures have been employed to assess the proposed model's suitability. After testing the major model assumptions, it is critical to assess the suggested model's effectiveness and predictive potential. Four statistical indices (i.e., $RMSE$, MAE , $MAPE$, and R^2) were utilized to examine the effectiveness of the suggested algorithm quantitatively, as shown in the equations below:

$$MSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (Y_i - \bar{Y}_i)^2} \quad (9)$$

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{Y_i - \bar{Y}_i}{Y_i} \right| \times 100 \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (Y_i - \bar{Y}_i)^2}{\sum_{i=1}^m (Y_i - \bar{Y})^2} \quad (11)$$

where Y_i represents the actual (measured) values of the overstrength ratio of short links, \bar{Y}_i represents the forecasted outcome, \bar{Y} represents the mean of the Y_i , and m represents the number of the datasets utilized. The model accuracy and performance will be enhanced if R^2 value approaches one as well as $RMSE$, MAE , and $MAPE$ measures approach zero.

4. Result and Discussion

4.1. Descriptive Statistics

Table 2 summarizes the surveyed experimental results of tested 110 short links. In addition, Table 3 and Figure 6 show the statistical information of collected databases. Table 3 includes several features such as b_f/t_f from 10 to 20.71 with an average of 13.51, d/t_w from 11.33 to 57.5 with an average of 36.66, $e/(M/V)$, from 0.33 to 1.69, A_f/A_w from 0.41 to 2.27 with an average of 1.86, $A_f f_{yflange}$ from 260 to 9882 kN with an average of 879 kN, $A_w f_{yweb}$ from 219.7 to 8524.3 kN with an average of 891.67 kN.

Table 2. Database description of short links.

Reference	No. of Tests	b_f/t_f	d/t_w	$e/(M/V)$	$f_{yflange}$ MPa	f_{yweb} MPa	V_{test} (kN)
Ji et al., 2015 [3]	12	12.9	40	0.58–0.97	319	228; 273	869–1130
Ji et al., 2016 [11]	2	10.6; 14.2	35	0.7–0.76	378; 396	228	838–926
McDaniel et al., 2003 [5]	2	10.6–13.3	33.9	0.59; 0.82	366	354	9363–9919
Volynkin et al., 2018 [46]	5	12–12.8	21.7–44.2	0.76–1.02	364; 455	364; 374	783–1034
Dusicka et al., 2010 [8]	5	11.8; 13.6	22–33.9	0.8; 0.82	223–503	242–503	1845–4348
Liu et al., 2017 [4]	11	10–13	21–35	1.12–1.6	366	354–362	373–668
Okazaki et al., 2005 [6]	11	11.5–18.3	22.1–56.8	1.04–1.49	319–362	382–404	585–1280
Okazaki, T. 2004 [7]	6	12.2	57.5	1.11	351.6	393	1007–1140
Bokurt and Topaya 2017 [12]	8	18–20.7	22.4–22.8	1.04–1.59	268–281	275–299	275–591
Bokurt and et al., 2019 [13]	6	18–20	22.2–29	1.26–1.59	272–357	276–343	288–573
Tong et al., 2018 [53]	4	12	17.9	1.25	461.2	463.4	720–1013
Mahmoudi et al., 2018 [54]	1	10	34	0.78	301	301	478
Hjelmstad et al., 1983 [45]	8	11.5; 15.6	43.4; 57	1.27–1.57	241.3; 285.4	711–914	600–1067
Dubina et al., 2008 [44]	24	12.25	38.7	0.65–1.3	221–315	221–315	270–420
Price, B. 2015 [43]	5	11.5; 16.5	23.8; 56.8	1.11; 1.23	353.7; 398.5	360; 403	433–1298
Total	110						

Table 3. Statistical description analysis of features.

Stander Statistics	Features					
	(b_f/t_f)	(d/t_w)	(A_f/A_w)	$(A_f f_{yflange})$	$(A_w f_{yweb})$	$e/(M/V)$
Mean	13.51	36.66	1.01	879.08	891.67	1.09
Standard Error	0.24	1.16	0.04	115.7	107.91	0.03
Median	12.24	38.71	0.86	608.74	664.32	1.1
Mode	12.24	38.71	0.86	803.88	550.24	0.87
Standard Deviation	2.53	12.18	0.43	1213.48	1131.79	0.28
Sample Variance	6.42	148.37	0.18	1,472,537	1,280,955	0.08
Kurtosis	0.56	−0.76	0.33	36.84	37.1	−0.65
Skewness	1.33	0.31	1.08	5.65	5.74	−0.15
Range	10.71	46.15	1.86	9622.04	8304.59	1.36
Minimum	10	11.33	0.41	259.96	219.73	0.33
Maximum	20.71	57.48	2.27	9882	8524.32	1.69
Sum	1486.02	4032.6	110.61	96698.71	98083.4	119.9
Count	110	110	110	110	110	110

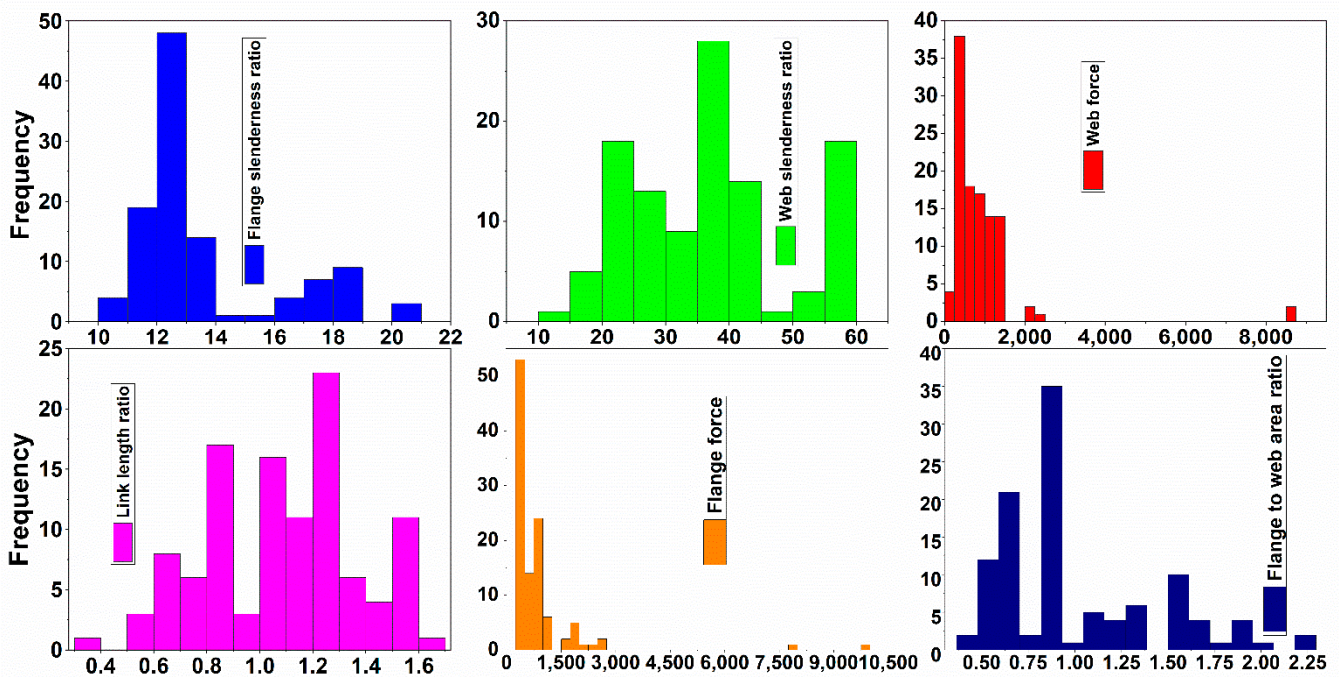


Figure 6. Database statistical description.

4.2. Correlation Matrix Analysis

Pearson's correlation among and in between selected features and the short links' shear strength was applied for the evaluation of the impact of these features, as shown in Figure 7. The relation's sign determines the trend of correlation between terms to investigate the effect of each item against every other item. The correlation factor ranges from +1 (strong positive relation) to -1 (strong negative relation); the correlation near zero indicates a weak relation. Figure 7 shows the correlation between flange force, web force, and link length ratio with the shear link strengths are 0.943, 0.965, and -0.904 , respectively. The results indicate that a strong relationship exists between these variables with the shear link strength. On the other hand, the correlation factors for web slenderness ratio, flange slenderness ratio, and flange to web area ratio are close to zero, which implies a minor effect of these parameters on the shear link strength.

Features	Flange Force	Web Force	Link Length Ratio	Web Slenderness Ratio	Flange Slenderness Ratio	Flange To Web Area Ratio	Shear Strength
Flange Force	1.000						
Web Force	0.935	1.000					
Link Length Ratio	-0.307	-0.227	1.000				
Web Slenderness Ratio	-0.137	0.122	-0.088	1.000			
Flange Slenderness Ratio	-0.116	-0.201	0.280	-0.401	1.000		
Flange To Web Area Ratio	0.124	-0.125	0.073	-0.684	0.502	1.000	
Shear Strength	0.943	0.965	-0.904	-0.096	-0.088	-0.074	1.000

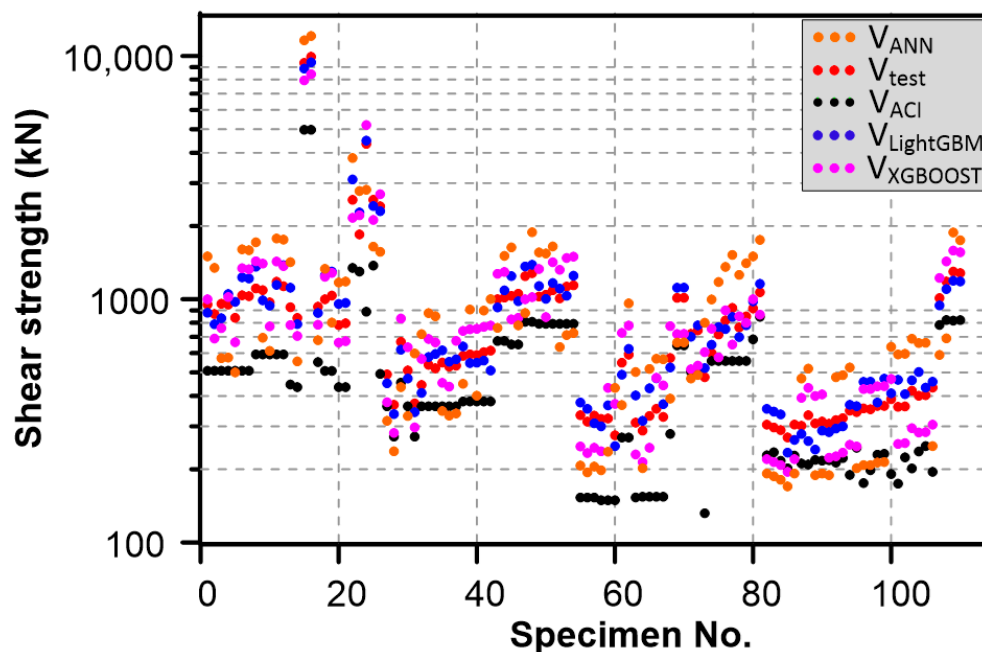
Figure 7. Correlation matrix analysis.

4.3. Performance of ML Algorithms

The goal of this research is to assess the efficiency of current *ML* methods (i.e., *ANN*, *XGBOOST*, and *LightGBM*) to predict the overstrength ratio of short links. Table 4 shows the five-evaluation comparison. The short links' shear strength predictions for various algorithms are displayed in Figure 8.

Table 4. Performance comparison for prediction of the ML models and AISC code.

Performance Comparison	Prediction Models			
	<i>LightGBM</i>	<i>XGBOOST</i>	<i>ANN</i>	<i>AISC Code</i>
<i>MAE</i>	92.0	196.5	378.0	397.9
<i>RMSE</i>	132.5	284.0	507.9	804.2
<i>MAPE</i>	11.7	24.1	35.8	39.2
R^2	0.99	0.96	0.90	0.75
Training Tim	7 s	9 s	14 s	

**Figure 8.** Model's prediction comparison with experimental strength.

The comparison outcomes present that *LightGBM* and *XGBOOST* have higher measures of R^2 value and lower *MAE*, *RMSE*, *MAPE* values than the traditional models (i.e., *ANN*, *AISC code*) in the short links' shear strength prediction. According to these metrics, the *LightGBM* and *XGBOOST* models surpassed other models. On the other hand, the results also indicate that *LightGBM* had outstanding prediction ability compared to *XGBOOST*. As a result, it generates a precise forecast in comparison to other proposed methodologies. The *LightGBM* was graded superior to the *XGBOOST*, as its assessment measures for *MAE* and *RMSE* were 104.5 and 151.5 points lower than the *XGBOOST*, respectively. Consequently, the *LightGBM* has been the most accurate model, as it had the minimum *MAE* and *RMSE* assessment scores. The *MAPE* was implemented to ensure that the *LightGBM* achieves the best overall accuracy when compared to other approaches. For *MAPE* measurement, *LightGBM*, *XGBOOST*, *ANN*, and *AISC code* achieved precisions of 11.7, 24.1, 35.8, and 39.2, respectively, as shown in Table 4. This means that the *XGBOOST*, *ANN*, and *AISC code* value of this measurement is much higher than *LightGBM*.

The result of R^2 is considered another realistic evidence that *LightGBM* is superior to all other approaches for determining the short links' shear strength, where, R^2 of *LightGBM* is greater than R^2 of *XGBOOST*, *ANN* and *AISC* subspecialty 3, 9, 24%. In comparison to other approaches, the *LightGBM* exhibits excellent forecasting accuracy. Moreover, the *LightGBM* model ran 2 s quicker than the *XGBOOST* model and 5 s quicker than the *ANN* model.

The forecasted results of the proposed illustrate that the *LightGBM* prediction values are very close to the experimental strength measures. Therefore, the better fit and slight deviation to the experimental values is the *LightGBM* model. The plots compare the predicted performance of the proposed models. As a result, the *LightGBM* is the most efficient and proficient model in the short links' shear strength prediction.

4.4. Features Importance Analysis

A better view of the model's features helps structural engineers effectively judge trends. Therefore, feature importance assessment has been performed using *LightGBM*, *XGBOOST*, and *ANN* models to figure out how important each variable is in predicting the overstrength ratio of short links. Thus, the feature score plot has been conducted to provide a relative score for each variable. Accordingly, the features' significance was in descending order: Web force, Link length ratio, Flange force, Web slenderness ratio, Flange slenderness ratio, and Flange to web area ratio, as shown in Figure 9.

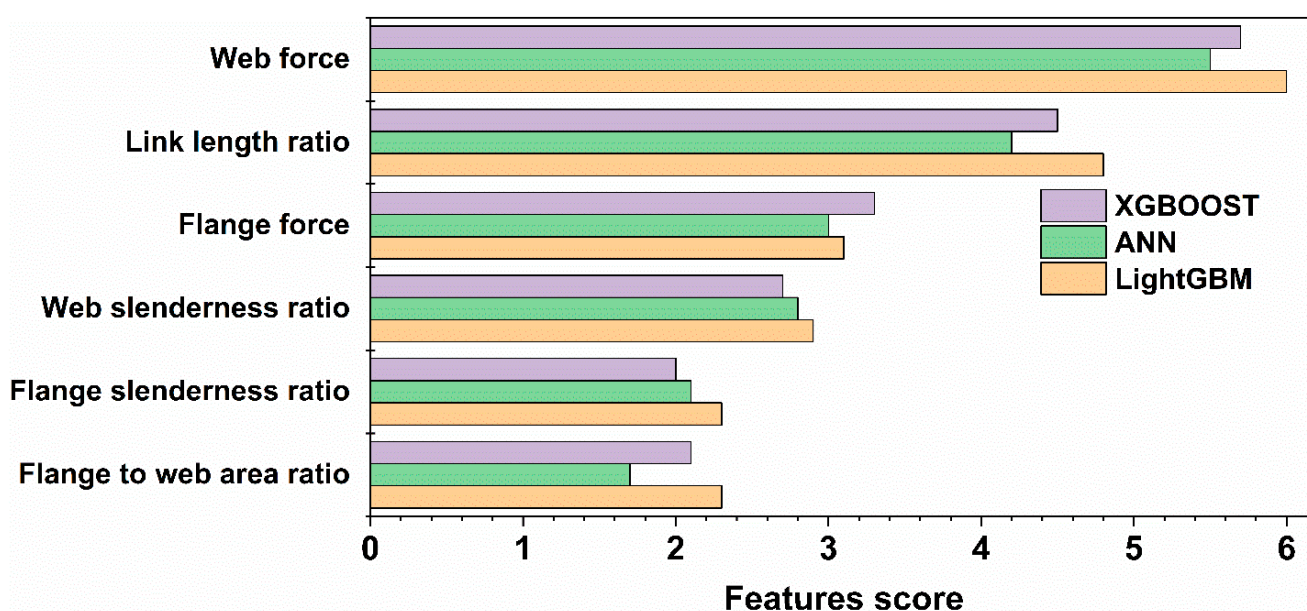


Figure 9. Representation of features importance.

The features' importance results are compatible with the literature where Liu et al., 2017 [4] noticed a significant impact of web slenderness and link length ratio on the short links' shear strength. The link length was critical in the steel links, Okazaki, T. 2004 [7]. Bozkurt and Topkaya 2017 [12] discovered that the overstrength ratio negatively relates to the link length ratio. Seven short links were tested, with numerous aspects such as link length ratio, stiffener spacing, and loading technique taken into account.

The overstrength ratio of experimental-to-*LightGBM* projected shear strength ($V_u/V_{LightGBM}$) and the experimental-to-AISC projected shear strength (V_u/V_p) are illustrated in Figures 10 and 11. Likewise, the AISC based overstrength ratio, the *LightGBM* demonstrated an excellent performance in the prediction of the shear link strength, where it is crucial to b_f/t_f , d/t_w , A_f/A_w , $A_f f_{yflange}$, $A_w f_{yweb}$, and $e/(M/V)$. The *LightGBM* predictions are flat and close to 1.0, which indicates that the *LightGBM* model is a comprehensive and competent algorithm for predicting the short links' shear strength.

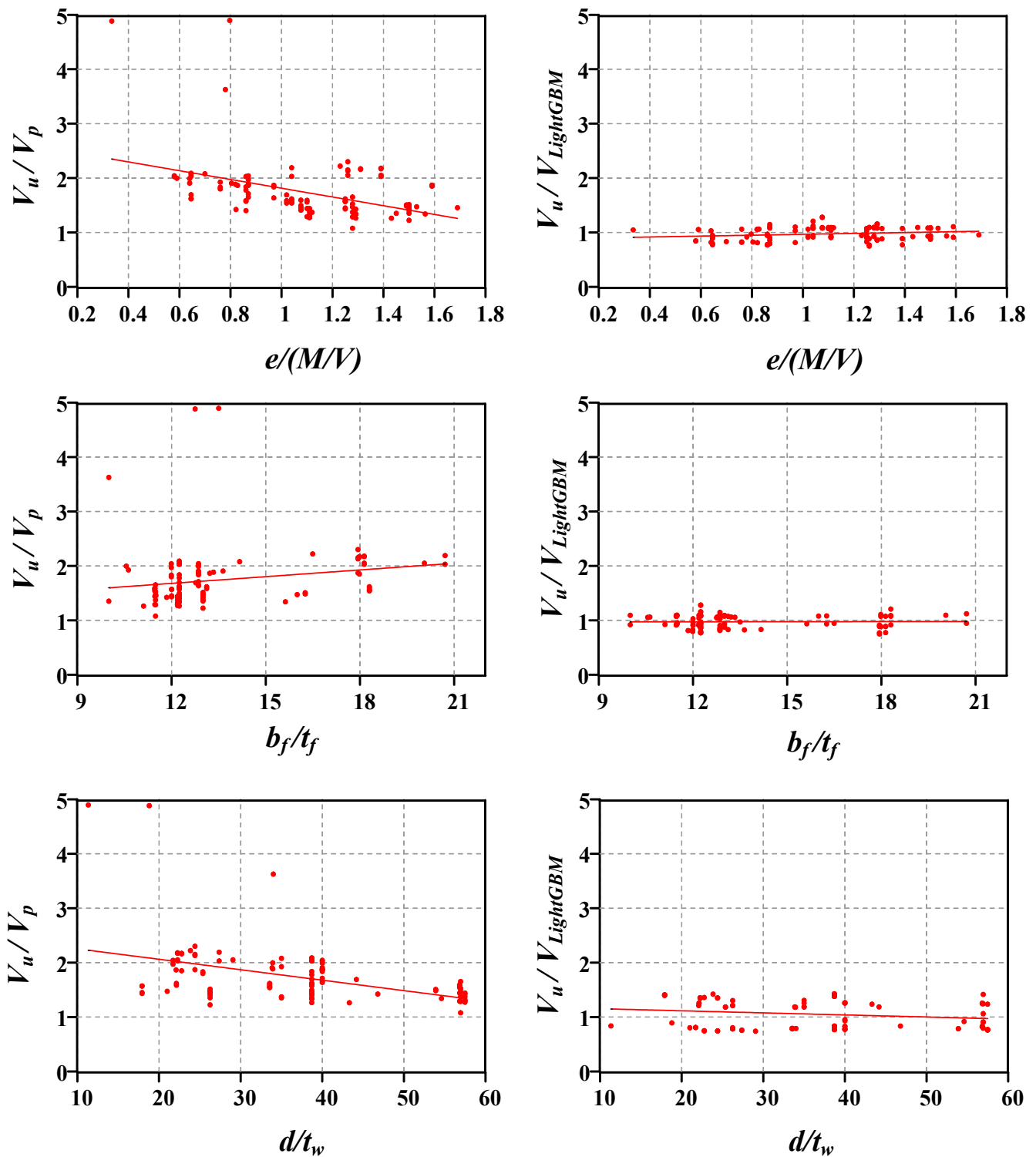


Figure 10. Distribution of overstrength ratio for the built-up sections.

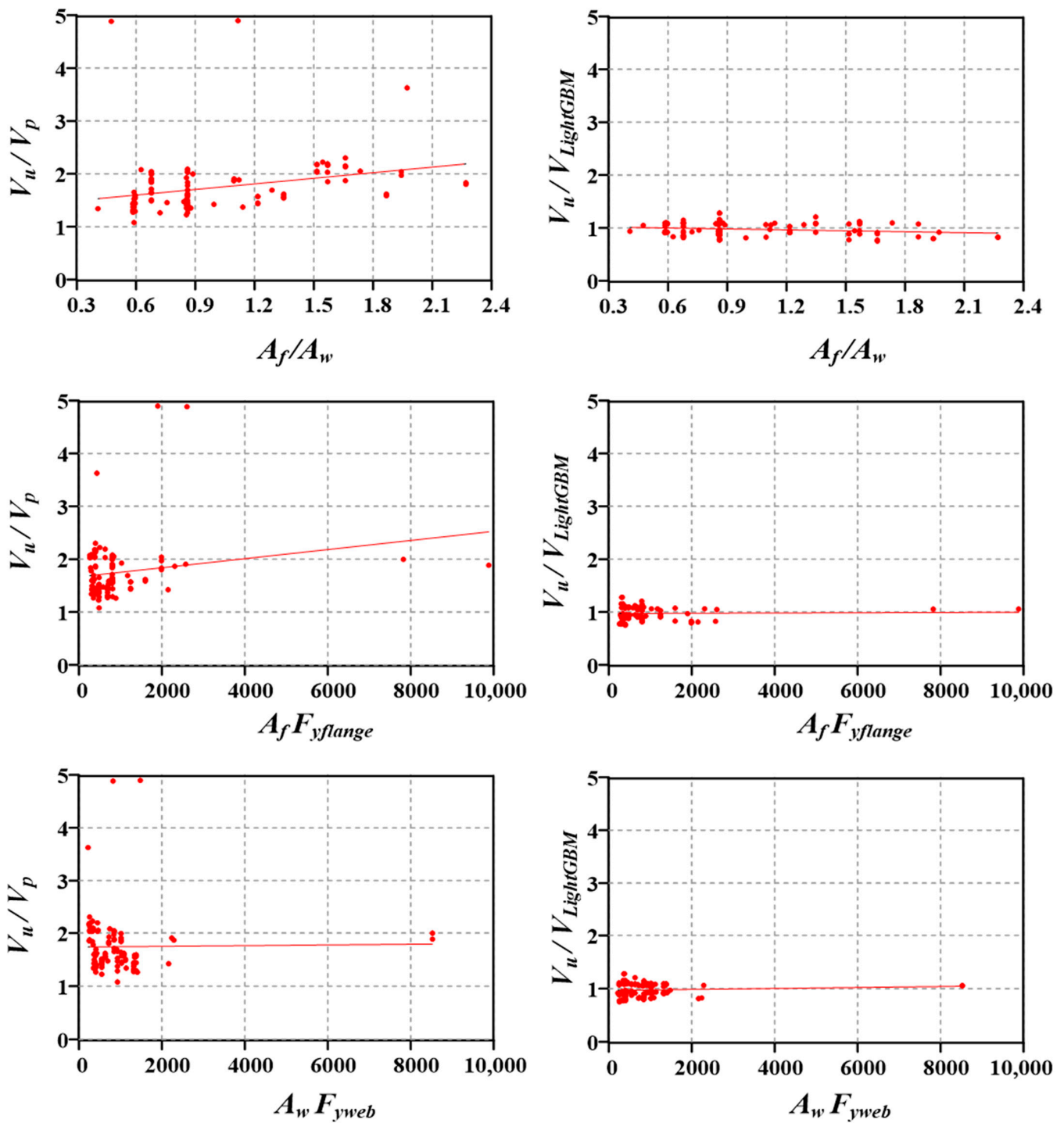


Figure 11. Distribution of overstrength ratio for the built-up sections.

Moreover, a comparison between the predicted overstrength ratio using *LightGBM* ($V_u/V_{LightGBM}$), AISC 2016 (V_u/V_p), Gene expression model (V_u/V_{GEP}), and FEM-based model ($V_{0.08}/V_y$) are presented in Figure 12. The average of the predicted overstrength ratio is 0.97, 1.11, 1.73, and 1.74 for the $V_u/V_{LightGBM}$, V_u/V_{GEP} , $V_{0.08}/V_y$, and V_u/V_p , respectively. The results revealed that the *LightGBM* is an excellent model in order to evaluate the short links' shear strength, while the AISC code equation is deficient to accurately estimate the shear link strength due to the fact the AISC code equation only considers the strength of web. This study revealed the significant effect of other variables such as the properties of flange

and the link length ratio. The machine learning algorithm (*LightGBM*) has successfully traced the contribution of elements other than the web on the short link shear strength.

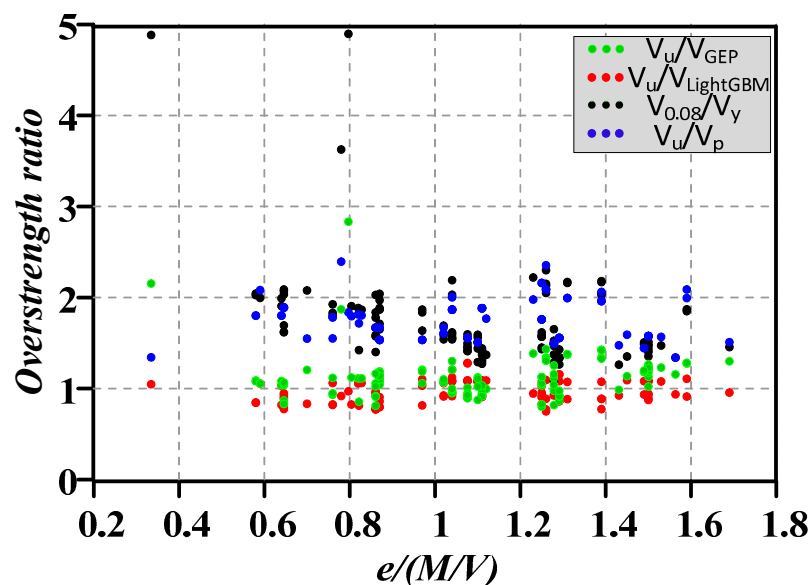


Figure 12. Performance of overstrength ratio vs. link length ratio for various models.

Technology is increasingly being used by policymakers to modify and construct policies. The existing research creates a strong *ML* framework for forecasting short links' shear strength that can be used as a universal model to mimic all relevant properties. This lays a solid platform for examining how various feature interconnections can help predict short links' shear strength. Furthermore, as modern *ML* methods improve, more advanced forecast models become available, allowing for the development of more effective and precise models for short-link shear strength prediction, which many construction industry stakeholders can employ. The researchers are optimistic in the suggested models' capabilities to supply decision-makers with more accurate projections to match available datasets as a beneficial prior acquaintance to power *ML*-based models, which is coherent with what is presently arising in the building business and construction projects.

In addition, as compared to other available models, created models were discovered to be convincing decision support aids in numerous sectors of the construction industry. The proposal could be a comprehensive, general, useful, and reliable prediction tool. Thus, the proposed methodology will play an important role in decreasing conflict among stakeholders in the steel construction sector, especially when decision-makers countenance massive problems and barriers indicating satisfactory short link shear strength that all contracting stakeholders agree on.

5. Conclusions

Although the machine learning algorithms were successfully implemented in the civil engineering applications, limited studies in the literature exploited the powerful capabilities of machine learning in estimating the shear link strength. The current study seeks to fill this gap with the help of the abundant short links experimental database in the literature to train, validate, and test sophisticated machine learning algorithms. Particularly, this research work examined the applications of the *ANN*, *XGBoost*, and *LightGBM* algorithms to enhance the prediction accuracy for the shear strength of short links using related features such as (b_f/t_f) , (d/t_w) , (A_f/A_w) , $(A_f f_{yflange})$, $(A_w f_{yweb})$, and $e/(M/V)$. The adopted forecasting process was based on 80% training datasets and 20% testing datasets to confirm the precision of the developed *ML* models. Performance metrics of the MAE, RMSE, MAPE, and R^2 under the 10-fold cross-validation process have been implemented to enhance the robustness and effectiveness of such models. These

measures reveal that the performance of the ML models set side by side to *AISC code* was arranged as follows: *LightGBM* > *XGBOOST* > *ANN* > *AISC code*. According to the importance of the features extracted from ML algorithms, Web force and Link length ratio were the most prominent variables in the prediction results of the overstrength ratio of short links. In addition, the predicted overstrength ratio using the *LightGBM* was compared to the available models in the literature, where the proficiency of developed models was reasonable. The analysis disclosed that the *LightGBM* has the least average predicted overstrength ratio compared to the GEP, FEM, or AISC-based models. For future research, a larger database can be adopted to demonstrate the adequacy of these models to predict the overstrength ratio of short links. The impact of other variables on the prediction accuracy needs to be adopted. Moreover, modern algorithms are required to improve results accuracy.

Author Contributions: Conceptualization, O.A. and G.A.; methodology, A.S.; software, A.S.; validation, G.A., O.A. and A.S.A.; formal analysis, G.A.; investigation, O.A.; resources, O.A. and A.S.A.; data curation, G.A.; writing—original draft preparation, A.S.; writing—review and editing, O.A.; visualization, A.S.; supervision, O.A.; project administration, G.A. and A.S.; funding acquisition, A.S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups Project under grant number (RGP. 2/178/43).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. AISC. *Seismic Provisions for Structural Steel Building*; ANSI/AISC 341-16; AISC: Chicago, IL, USA, 2016.
2. Engelhardt, M.D.; Popov, E.P. Experimental Performance of Long Links in Eccentrically Braced Frames. *J. Struct. Eng.* **1992**, *118*, 3067–3088. [[CrossRef](#)]
3. Ji, X.; Wang, Y.; Ma, Q.; Okazaki, T. Cyclic Behavior of Very Short Steel Shear Links. *J. Struct. Eng.* **2015**, *142*, 04015114. [[CrossRef](#)]
4. Liu, X.-G.; Fan, J.-S.; Liu, Y.-F.; Yue, Q.-R.; Nie, J.-G. Experimental research of replaceable Q345GJ steel shear links considering cyclic buckling and plastic overstrength. *J. Constr. Steel Res.* **2017**, *134*, 160–179. [[CrossRef](#)]
5. McDaniel, C.C.; Uang, C.-M.; Seible, F. Cyclic Testing of Built-Up Steel Shear Links for the New Bay Bridge. *J. Struct. Eng.* **2003**, *129*, 801–809. [[CrossRef](#)]
6. Okazaki, T.; Arce, G.; Ryu, H.-C.; Engelhardt, M.D. Experimental Study of Local Buckling, Overstrength, and Fracture of Links in Eccentrically Braced Frames. *J. Struct. Eng.* **2005**, *131*, 1526–1535. [[CrossRef](#)]
7. Okazaki, T. *Seismic Performance of Link-To Column Connections in Steel Eccentrically Braced Frames*; The University of Texas at Austin: Austin, TX, USA, 2004.
8. Dusicka, P.; Itani, A.M.; Buckle, I.G. Cyclic Behavior of Shear Links of Various Grades of Plate Steel. *J. Struct. Eng.* **2010**, *136*, 370–378. [[CrossRef](#)]
9. AISC. *Seismic Provisions for Structural Steel Buildings*; ANSI/AISC 341-02; AISC: Chicago, IL, USA, 2002.
10. AISC. *Specification for Structural Steel Buildings*; ANSI/AISC 360-10; AISC: Chicago, IL, USA, 2010.
11. Ji, X.; Wang, Y.; Ma, Q.; Okazaki, T. Cyclic Behavior of Replaceable Steel Coupling Beams. *J. Struct. Eng.* **2016**, *143*, 04016169. [[CrossRef](#)]
12. Bozkurt, M.B.; Topkaya, C. Replaceable links with direct brace attachments for eccentrically braced frames: Replaceable Links with Direct Brace Attachments for EBF. *Earthq. Eng. Struct. Dyn.* **2017**, *46*, 2121–2139. [[CrossRef](#)]
13. Bozkurt, M.B.; Kazemzadeh Azad, S.; Topkaya, C. Development of detachable replaceable links for eccentrically braced frames. *Earthq. Eng. Struct. Dyn.* **2019**, *48*, 1134–1155. [[CrossRef](#)]
14. Chao, S.-H.; Khandelwal, K.; El-Tawil, S. Ductile Web Fracture Initiation in Steel Shear Links. *J. Struct. Eng.* **2006**, *132*, 1192–1200. [[CrossRef](#)]
15. Della Corte, G.; D’Aniello, M.; Landolfo, R. Analytical and numerical study of plastic overstrength of shear links. *J. Constr. Steel Res.* **2013**, *82*, 19–32. [[CrossRef](#)]
16. Hong, J.-K.; Uang, C.-M.; Okazaki, T.; Engelhardt, M.D. Link-to-Column Connection with Supplemental Web Doublers in Eccentrically Braced Frames. *J. Struct. Eng.* **2015**, *141*, 04014200. [[CrossRef](#)]
17. Hu, S.; Xiong, J.; Zhou, Q.; Lin, Z. Analytical and Numerical Investigation of Overstrength Factors for Very Short Shear Links in EBFs. *KSCE J. Civ. Eng.* **2018**, *22*, 4473–4482. [[CrossRef](#)]

18. Liu, X.-G.; Fan, J.-S.; Liu, Y.-F.; Zheng, M.-Z.; Nie, J.-G. Theoretical research into cyclic web buckling and plastic overstrength of shear links. *Thin Walled Struct.* **2020**, *152*, 106644. [[CrossRef](#)]
19. Ohsaki, M.; Nakajima, T. Optimization of link member of eccentrically braced frames for maximum energy dissipation. *J. Constr. Steel Res.* **2012**, *75*, 38–44. [[CrossRef](#)]
20. Yin, W.-H.; Sun, F.-F.; Jin, H.-J.; Hu, D.-Z. Experimental and analytical study on plastic overstrength of shear links covering the full range of length ratio. *Eng. Struct.* **2020**, *220*, 110961. [[CrossRef](#)]
21. Song, H.; Ahmad, A.; Farooq, F.; Ostrowski, K.A.; Maslak, M.; Czarnecki, S.; Aslam, F. Predicting the compressive strength of concrete with fly ash admixture using machine learning algorithms. *Constr. Build. Mater.* **2021**, *308*, 125021. [[CrossRef](#)]
22. Tarawneh, A.; Almasabha, G.; Murad, Y. ColumnsNet: Neural Network Model for Constructing Interaction Diagrams and Slenderness Limit for FRP-RC Columns. *J. Struct. Eng.* **2022**, *148*, 04022089. [[CrossRef](#)]
23. Saleh, E.; Tarawneh, A.; Naser, M.; Abedi, M.; Almasabha, G. You only design once (YODO): Gaussian Process-Batch Bayesian optimization framework for mixture design of ultra high performance concrete. *Constr. Build. Mater.* **2022**, *330*, 127270. [[CrossRef](#)]
24. Almasabha, G.; Tarawneh, A.; Saleh, E.; Alajarmeh, O. Data-Driven Flexural Stiffness Model of FRP-Reinforced Concrete Slender Columns. *J. Compos. Constr.* **2022**, *26*, 04022024. [[CrossRef](#)]
25. Tarawneh, A.; Almasabha, G.; Alawadi, R.; Tarawneh, M. Innovative and Reliable Model for Shear Strength of Steel Fibers Reinforced Concrete Beams. *Structures* **2021**, *32*, 1015–1025. [[CrossRef](#)]
26. Alshboul, O.; Alzubaidi, M.A.; Mamlook, R.E.A.; Almasabha, G.; Almuflih, A.S.; Shehadeh, A. Forecasting Liquidated Damages via Machine Learning-Based Modified Regression Models for Highway Construction Projects. *Sustainability* **2022**, *14*, 5835. [[CrossRef](#)]
27. Alshboul, O.; Shehadeh, A.; Tatari, O.; Almasabha, G.; Saleh, E. Multiobjective and multivariable optimization for earthmoving equipment. *J. Facil. Manag.* **2022**. [[CrossRef](#)]
28. Shehadeh, A.; Alshboul, O.; Tatari, O.; Alzubaidi, M.A.; Hamed El-Sayed Salama, A. Selection of heavy machinery for earthwork activities: A multi-objective optimization approach using a genetic algorithm. *Alex. Eng. J.* **2022**, *61*, 7555–7569. [[CrossRef](#)]
29. Alshboul, O.; Shehadeh, A.; Hamedat, O. Development of integrated asset management model for highway facilities based on risk evaluation. *Int. J. Constr. Manag.* **2021**, 1–10. [[CrossRef](#)]
30. Shehadeh, A.; Alshboul, O.; Hamedat, O. A Gaussian mixture model evaluation of construction companies' business acceptance capabilities in performing construction and maintenance activities during COVID-19 pandemic. *Int. J. Manag. Sci. Eng. Manag.* **2022**, *17*, 112–122. [[CrossRef](#)]
31. Alshboul, O.; Shehadeh, A.; Hamedat, O. Governmental Investment Impacts on the Construction Sector Considering the Liquidity Trap. *J. Manag. Eng.* **2022**, *38*, 04021099. [[CrossRef](#)]
32. Shehadeh, A.; Alshboul, O.; Hamedat, O. Risk Assessment Model for Optimal Gain-Pain Share Ratio in Target Cost Contract for Construction Projects. *J. Constr. Eng. Manag.* **2022**, *148*, 04021197. [[CrossRef](#)]
33. Alshboul, O.; Shehadeh, A.; Almasabha, G.; Almuflih, A.S. Extreme Gradient Boosting-Based Machine Learning Approach for Green Building Cost Prediction. *Sustainability* **2022**, *14*, 6651. [[CrossRef](#)]
34. Almasabha, G. Gene expression model to estimate the overstrength ratio of short links. *Structures* **2022**, *37*, 528–535. [[CrossRef](#)]
35. Alshboul, O.; Shehadeh, A.; Al-Kasasbeh, M.; Al Mamlook, R.E.; Halalshah, N.; Alkasasbeh, M. Deep and machine learning approaches for forecasting the residual value of heavy construction equipment: A management decision support model. *Eng. Constr. Archit. Manag.* **2021**. [[CrossRef](#)]
36. Shehadeh, A.; Alshboul, O.; Al Mamlook, R.E.; Hamedat, O. Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression. *Autom. Constr.* **2021**, *129*, 103827. [[CrossRef](#)]
37. Cevik, A. Genetic programming based formulation of rotation capacity of wide flange beams. *J. Constr. Steel Res.* **2007**, *63*, 884–893. [[CrossRef](#)]
38. Fonseca, E.T.; da Vellasco, P.C.G.S.; de Andrade, S.A.L.; Vellasco, M.M.B.R. Neural network evaluation of steel beam patch load capacity. *Adv. Eng. Softw.* **2003**, *34*, 763–772. [[CrossRef](#)]
39. Güneyisi, E.M.; D'Aniello, M.; Landolfo, R.; Mermerdaş, K. A novel formulation of the flexural overstrength factor for steel beams. *J. Constr. Steel Res.* **2013**, *90*, 60–71. [[CrossRef](#)]
40. Fan, J.; Wang, X.; Wu, L.; Zhou, H.; Zhang, F.; Yu, X.; Lu, X.; Xiang, Y. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Convers. Manag.* **2018**, *164*, 102–111. [[CrossRef](#)]
41. Zhang, M.; Xiang, F.; Liu, Z. Short-term traffic flow prediction based on combination model of XGBoost-LightGBM. In Proceedings of the 2018 International Conference on Sensor Networks and Signal Processing (SNSP), Xi'an, China, 28–31 October 2018; pp. 322–327.
42. Pathy, A.S.; Meher, B.P. Predicting algal biochar yield using eXtreme Gradient Boosting (XGB) algorithm of machine learning methods. *Algal Res.* **2020**, *50*, 102006. [[CrossRef](#)]
43. Price, B. Investigation on Innovative Shear Link Configurations and Optimal Design for Earthquake Resistant Steel Eccentrically Braced Frames. Master's Thesis, University of Texas at Arlington, Arlington, TX, USA, 2015.
44. Dubina, D.; Stratan, A.; Dinu, F. Dual high-strength steel eccentrically braced frames with removable links. *Earthq. Eng. Struct. Dyn.* **2008**, *37*, 1703–1720. [[CrossRef](#)]

45. Hjelmstad, K.D.; Popov, E.P. Cyclic Behavior and Design of Link Beams. *J. Struct. Eng.* **1983**, *109*, 2387–2403. [[CrossRef](#)]
46. Volynkin, D.; Dusicka, P.; Clifton, G.C. Intermediate Web Stiffener Spacing Evaluation for Shear Links. *J. Struct. Eng.* **2018**, *145*, 04018257. [[CrossRef](#)]
47. Yang, X. Artificial neural networks. In *Handbook of Research on Geoinformatics*; IGI Global: Hershey, PA, USA, 2009; pp. 122–128.
48. Kim, M.; Jung, S.; Kang, J. Artificial Neural Network-Based Residential Energy Consumption Prediction Models Considering Residential Building Information and User Features in South Korea. *Sustainability* **2020**, *12*, 109. [[CrossRef](#)]
49. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [[CrossRef](#)]
50. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **2001**, *29*, 11891232. [[CrossRef](#)]
51. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. *Adv. Neur. Infor. Process. Sys.* **2017**, *30*, 3146–3154.
52. Zeng, H.; Yang, C.; Zhang, H.; Wu, Z.; Zhang, J.; Dai, G.; Babiloni, F.; Kong, W. A lightGBM-based EEG analysis method for driver mental states classification. *Comput. Intell. Neurosci.* **2019**, *2019*, 3761203. [[CrossRef](#)]
53. Tong, L.; Zhang, Y.; Zhang, L.; Liu, H.; Zhang, Z.; Li, R. Ductility and energy dissipation behavior of G20Mn5QT cast steel shear link beams under cyclic loading. *J. Constr. Steel Res.* **2018**, *149*, 64–77. [[CrossRef](#)]
54. Mahmoudi, F.; Dolatshahi, K.M.; Mahsuli, M.; Nikoukalam, M.T.; Shahmohammadi, A. Experimental study of steel moment resisting frames with shear link. *J. Constr. Steel Res.* **2018**, *154*, 197–208. [[CrossRef](#)]