Challenge Journal of STRUCTURAL MECHANICS

Research Article

Data-driven prediction of initial rotational stiffness in beam-to-upright connections of steel pallet racks

Casim Yazici [a], F. Javier Dominguez-Gutierrez [b,*]

[a] *Department of Construction, Ağrı İbrahim Çeçen University, 04400 Ağrı, Türkiye*
[b] *NOMATEN Centre of Excellence, National Centre for Nuclear Research, ul. Andrzeja Soltana 7, 05-400 Świerk, Poland*

ABSTRACT

The structural performance of steel pallet rack systems is largely governed by the rotational behavior of beam-to-upright connections. Although design practice often idealizes these joints as rigid or pinned, experimental evidence shows that most exhibit a semi-rigid response. Realistic structural analysis therefore requires a reliable estimate of the initial rotational stiffness, k_0, which controls the moment–rotation relationship and affects internal force distribution and frame stability. This study proposes a data-driven framework for predicting k_0 directly from the geometric and mechanical properties of uprights, beams, and connectors. The model incorporates key sectional parameters and component yield strengths and is trained on a consolidated database compiled from experimentally validated literature. A comparative assessment of regression techniques shows that ensemble-based machine-learning models, particularly the Extra Trees regressor, provide the highest predictive accuracy. Feature importance and SHAP analyses indicate that connection stiffness is primarily governed by geometric parameters controlling load transfer and deformation mechanisms, while material strength plays a secondary role. The proposed approach eliminates the need for extensive experimental testing and supports realistic semi-rigid modelling and more economical pallet rack design.

ARTICLE INFO

Article history:
Received – January 15, 2026
Revision requested – February 13, 2026
Revision received – February 24, 2026
Accepted – March 3, 2026

Keywords:
Pallet rack systems
Cold-formed steel
Semi-rigid connections
Beam-to-upright joint
Rotational stiffness
Machine learning
Data-driven modelling

Citation: Yazici C, Dominguez-Gutierrez FJ (2026). Data-driven prediction of initial rotational stiffness in beam-to-upright connections of steel pallet. *Challenge Journal of Structural Mechanics*, 12(2), 73–86.

1. Introduction

Steel storage rack structures assembled from cold-formed, thin-walled members are pervasive in warehousing and retail logistics owing to their modularity, load efficiency, and cost effectiveness. In unbraced down-aisle frames, the global lateral response is governed by the semi-rigid behavior of beam-to-upright joints, typically realized through boltless hook-in beam-end connectors or through bolted “speed-lock” variants. The joint’s moment–rotation (M–θ) relationship and, in particular, the initial rotational stiffness k_0 in kN·m/rad decisively influence sway stability, second-order effects, and the reliability of design-by-analysis. Recognizing the diversity of commercial connector geometries and slot patterns across manufacturers, modern specifications endorse experimental characterization via standardized cantilever and portal setups to obtain consistent M–θ curves for analysis and design (Prabha et al. 2010; Mohan and Vishnu 2013; Zhao et al. 2018; Shah et al. 2016; Yin et al. 2016; Dai et al. 2018; Ślęczka and Kozłowski 2007; Shariati et al. 2018; Mohan et al. 2015; Shah et al. 2017). The accumulated experimental literature demonstrates that stiffness and strength depend sensitively on connector detailing and cross-sectional proportions of the participating members. Beam height

* *Corresponding author*. E-mail address: javier.dominguez@ncbj.gov.pl (F. J. Dominguez-Gutierrez)
ISSN: 2149-8024 / DOI: https://doi.org/10.20528/cjsmec.2026.02.002

(*hb*) and connector height (*hc*) have been shown to enhance rotational stiffness through increased lever arm and bearing interface, while the number of tabs (*tn*) and the thicknesses of the upright (*tu*), beam (*tb*), and connector (*tc*) condition the onset of local mechanisms such as tab tear-out, web cracking around perforations, and localized/distortional buckling (Prabha et al. 2010; Mohan and Vishnu 2013; Zhao et al. 2018; Shah et al. 2016; Yin et al. 2016; Dai et al. 2018; Zhao et al. 2014; Yazici et al. 2023; Shah et al. 2017). Under cyclic actions, boltless joints exhibit strongly pinched hysteresis with progressive stiffness degradation, confirming the semi-rigid, partial-strength nature of these connections and underscoring the need to model both backbone and reloading branches for seismic assessments (Zhao et al. 2018; Dai et al. 2018; Mohan et al. 2015). When bolts or welds supplement the hook-in details, ductility and energy dissipation improve markedly, although the increase in k_0 is often moderate once geometry is held constant (Yin et al. 2016; Dai et al. 2018; Mohan et al. 2015). These findings collectively motivate joint models that respect experimentally observed degradation and pinching while preserving a transparent link to measurable geometric and material descriptors.

Test procedures have evolved toward consolidation and comparability. Whereas earlier practice admitted several alternative fixtures, recent European practice emphasizes a single cantilever protocol for stiffness identification, facilitating direct comparison of results across product lines and laboratories (Prabha et al. 2010; Mohan and Vishnu 2013; Shariati et al. 2018). Portal frame arrangements remain useful for system-level studies but may obscure idiosyncratic single-joint responses through averaging effects. In parallel, definitional choices for stiffness have attracted attention. The initial-tangent method, the slope-to-half-ultimate method, and the equal-area (energy) method have all been deployed in the literature; comparative analyses caution that the raw initial tangent may bias k_0 upward in the presence of measurement noise and early nonlinearity, whereas energy-based estimates tend to be more robust across geometries and loading histories (Prabha et al. 2010; Mohan and Vishnu 2013; Shah et al. 2016; Zhao et al. 2014). The choice of definition has immediate implications for both experimental reporting and for validating computational models that seek to reproduce the shape of M–θ curves. Mechanistic studies using the component method and detailed Finite Element (FE) analyses complement the test evidence and help interpret where rotations concentrate. These models decompose contributions from tab–slot bearing and shear, connector plate bending, upright web distortion, and contact-friction interfaces. They also highlight that the instantaneous rotation center is not stationary but migrates with load progression, modulating the measured stiffness and the partitioning of deformations among subcomponents (Shah et al. 2016; Santamaria et al. 2024; Rosli et al. 2025; Yazici et al. 2023). Such insights illuminate why apparently similar connectors can yield distinct stiffness values once perforation geometry, tab spacing, and local plate slenderness differ.

Despite the maturity of the physical testing and modelling toolbox, there remains a practical challenge at the design stage. The space of admissible configurations spanning column flange width (*w*) and height (*b*), beam width (*bw*) and height (*hb*), connector height (*hc*), thicknesses (*tu, tb, tc*), tab number (*tn*), beam position (*bp*), and material yields of the upright, beam, and connector (*yu, yb, yc*) is combinatorial rich. Exhaustive prototype testing or high-fidelity FE exploration over this space is expensive and time-consuming for each new product family (Shah et al. 2016; Santamaria et al. 2024; Rosli et al. 2025; Yazici et al. 2023). Moreover, the predictive quantity of direct interest for global frame analysis is k_0 at modest rotations, not only ultimate resistance or ductility. Taken together, these considerations motivate complementary, data-driven estimators that map readily available design-stage descriptors to k_0 while remaining faithful to the patterns established by experiments and FE studies (Prabha et al. 2010; Mohan and Vishnu 2013; Zhao et al. 2018; Shah et al. 2016; Yin et al. 2016; Dai et al. 2018; Shah et al. 2015; Zhao et al. 2014; Shariati et al. 2018; Mohan et al. 2015; Yazici et al. 2023; Shah et al. 2017). Recent rack-connection studies have begun to explore data-driven prediction in tandem with classical mechanics-based models. Evidence suggests that supervised learning can capture nonlinear interactions among geometry, thickness, and yield strength to forecast stiffness and capacity with competitive accuracy relative to narrowly targeted test programs or single-configuration FE analyses (Santamaria et al. 2024; Rosli et al. 2025). Yet many prior data-centric efforts rely on restricted feature sets often limited to a few geometric quantities such as *hb, hc,* and *tu* and consequently do not encode scale-free proportions or positional effects that engineers naturally consider in preliminary design. This limitation motivates the inclusion of mechanics-informed features that reflect invariances and relative scales intrinsic to thin-walled behavior. In particular, dimensionless ratios such as *bw/w* and *hb/b* convey cross-sectional proportion and lever-arm effects; *tu/tb* reflects relative plate slenderness and stiffness between the primary members; and a normalized beam placement *bp/w* captures how eccentricity and bearing geometry influence the contact mechanics within the connector-upright assembly. When these features are paired with yield-strength descriptors (*yu, yb, yc*) and subjected to input standardization, they provide a more uniform numerical landscape for learning algorithms while preserving physically interpretable sensitivities that can be cross-checked against experimental trends documented in the literature (Prabha et al. 2010; Mohan and Vishnu 2013; Zhao et al. 2018; Shah et al. 2016; Yin et al. 2016; Dai et al. 2018; Zhao et al. 2014; Shariati et al. 2018; Mohan et al. 2015; Yazici et al. 2023; Shah et al. 2017).

In this study, a machine learning framework was developed by compiling 151 experimental data points from 15 published studies on beam-to-upright connections. Although the dataset is relatively modest for machine learning applications, rigorous cross-validation and independent test evaluation were employed to mitigate overfitting. Nevertheless, expanding the database with additional experimental results would further improve model generalization and robustness. By focusing on key section and connection properties reported in the literature, the predictive models provide rapid and reliable es-

timation of connector rotational stiffness in beam-to-upright joints. This enables researchers and practitioners to anticipate connection behavior prior to extensive testing, streamline experimental design, and support code-based design checks when direct test data are not available. Beyond predictive capability, the framework highlights the most influential parameters governing connector rotational stiffness, thereby allowing practitioners in the private sector to design sections more efficiently and enabling academics to plan targeted experimental programs guided by the model's outputs. In doing so, the methodology not only reduces the substantial costs associated with large-scale testing but also contributes a novel data-driven pathway for advancing both industrial applications and academic research on storage rack connections. The machine learning (ML) model developed in this study aims to predict the rotational stiffness of beam–column connections in storage rack systems. For this purpose, a comprehensive dataset was compiled using the section and connection parameters together with the corresponding experimentally obtained rotational stiffness values reported in the literature (Prabha et al. 2010; Mohan and Vishnu 2013; Zhao et al. 2018; Shah et al. 2016; Yin et al. 2016; Dai et al. 2018; Ślęczka and Kozłowski 2007; Santamaria et al. 2024; Shah et al. 2015; Zhao et al. 2014; Shariati et al. 2018; Mohan et al. 2015; Rosli et al. 2025; Yazici et al. 2023; Shah et al. 2017). This dataset was subsequently used in the training and validation stages of the proposed ML model.

2. Methods

The present study develops a supervised machine-learning model to estimate the initial rotational stiffness k_0 of beam-to-upright connections directly from the full suite of design variables available at the concept stage, namely *w, b, bw, hb, hc, tu, tb, tc, tn, bp, yu, yb,* and *yc*. The model is trained on a consolidated database curated from published experimental and numerical investigations of boltless and bolted rack joints, including monotonic and cyclic programs that characterize backbone curves, hysteretic pinching, and stiffness degradation (Prabha et al. 2010; Mohan and Vishnu 2013; Zhao et al. 2018; Shah et al. 2016; Yin et al. 2016; Dai et al. 2018; Ślęczka and Kozłowski 2007; Zhao et al. 2014; Shariati et al. 2018; Mohan et al. 2015; Yazici et al. 2023; Shah et al. 2017). Feature engineering leverages mechanically motivated, scale-free ratios (*bw/w, hb/b, tu/tb*) and the normalized placement parameter (*bp/w*), alongside yield-strength indicators, and inputs are standardized to mitigate scale disparities among dimensions, thicknesses, and stresses. The objective is to furnish fast, reproducible k_0 estimates suitable for preliminary design and for global analyses that require realistic semi-rigid joint idealizations, while retaining interpretability so that the learned parameter influences can be compared against established findings for example, the consistent positive association of beam depth and tab count with stiffness, and the limited incremental k_0 gains attributable to speed-lock bolting under fixed geometry (Zhao et al. 2018; Yin et al. 2016; Dai et al. 2018; Zhao et al. 2014; Mohan et al. 2015; Shah et al. 2017).

2.1. Geometric and mechanical parameters of the beam-to-upright connection

The geometric configuration of the investigated beam-to-upright connection and the associated input parameters are summarized in Fig. 1. The upright cross-section is included in this figure, where the flange width (*w*), overall section height (*b*), and upright thickness (*tu*) are defined. These quantities primarily influence the local stiffness of the perforated zone and, consequently, the resistance and deformation patterns developed during tab engagement. The beam section is also provided in Fig. 1. The beam width (*bw*), beam height (*hb*), and beam thickness (*tb*) are introduced as the principal geometric descriptors of the cold-formed member. Variations in these parameters affect the flexural rigidity of the beam and contribute to the global moment–rotation response of the connection. The connector and tab arrangement are further illustrated in Fig. 1, defining the connector height (*hc*), connector thickness (*tc*), and the number of engaged tabs (*tn*). In boltless rack connections, the tab engagement mechanism governs the dominant deformation modes, which typically involve tab bending, localized bearing at the upright perforations, and associated local distortions. For this reason, the connector geometry and the tab number were treated as key parameters within the adopted variable set. The assembled configuration is likewise presented in Fig. 1, emphasizing the position of the beam relative to the upright and connector. The beam position (*bp*) was incorporated to capture its effect on the internal force transfer path and the resulting rotational response. In addition to geometric descriptors, the material response of the connection components was represented through the yield strengths of the upright (*yu*), beam (*yb*), and connector (*yc*). Together with the geometric parameters defined above, these variables form the input space of the predictive models considered in this study. The connection rotational stiffness, k_0, was selected as the target output parameter and was determined in accordance with the procedures specified in EN 15512 (EN 15512 2022; Yazici et al. 2026; Özkal and Yazici 2024). This choice was made to ensure consistency across the multi-source database, since the raw initial tangent stiffness at the origin is highly sensitive to early slip/seating effects and measurement noise, and is not reported in a uniform manner for many published tests. The adopted EN 15512 definition provides a standardized and practically relevant stiffness metric for semi-rigid rack analysis and enables reliable aggregation of results from heterogeneous experimental programs.

2.2. Calculation of rotational stiffness for beam-to-upright connections

In this study, the rotational stiffness (k_0) of boltless beam-to-upright connections was determined in strict accordance with the experimental procedures specified in EN 15512 (EN 15512 2022). This parameter governs the semi-rigid response of the joint, for which the load transfer mechanism is based solely on the mechanical interlocking of the connector tabs with the perforations of the upright (EN 15512 2022; Yazici et al. 2026). The bending moment (*M*) acting on the connection was com-

puted from the vertical load (F) applied by the hydraulic actuator and the lever arm (a), which was kept constant at 400 mm from the face of the upright (EN 15512 2022). Accordingly, the moment was evaluated as: $M = F \times a$. The connection rotation (θ) was obtained from the relative horizontal displacements recorded by Linear Variable Differential Transformers (LVDTs). For this purpose, two LVDTs were attached to the upper part of the connector and one to the lower part (EN 15512 2022; Yazici et al. 2026; Özkal and Yazici 2024). The rotation was then calculated using:

$$\theta = \frac{\delta_{upper} - \delta_{lower}}{d} \quad (1)$$

where δ_{upper} and δ_{lower} represent the horizontal displacements measured at the upper and lower levels of the connector, respectively, and d denotes the vertical gauge length between these two measurement points (EN 15512 2022). The resulting rotation primarily captures the deformation of the connector tabs and the local distortion of the upright perforations, which are associated with Failure Mode (EN 15512 2022; Yazici et al. 2026). The overall test setup and a typical boltless beam-to-upright connection, including the lever arm a, the vertical gauge length d, and the positions of the three LVDTs used to evaluate the connection rotation, are illustrated in Fig. 2.

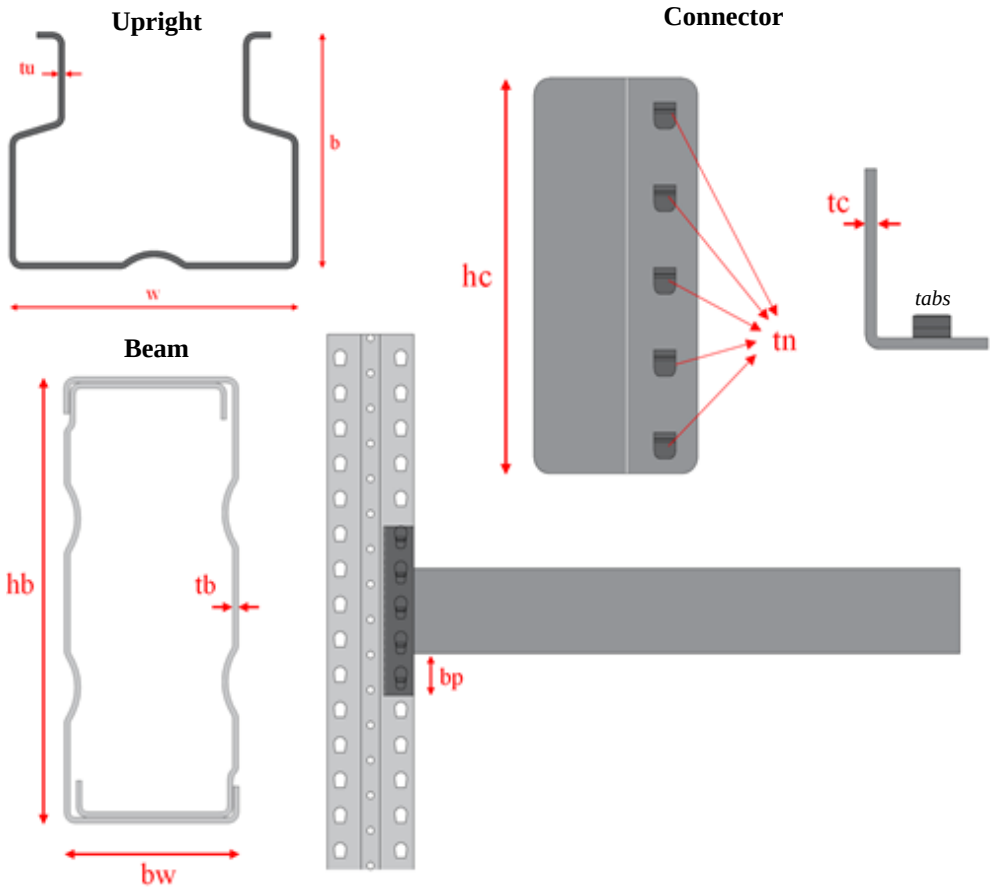

Fig. 1. Geometric parameters of the upright section used in the beam-to-upright connection, geometric parameters of the cold-formed beam section, connector geometry and tab configuration defining the number of engaged tabs, and beam-to-upright connection assembly showing the beam position relative to the upright and connector.

The initial rotational stiffness (k_0) was defined as the slope of the ascending branch of the moment–rotation response (EN 15512 2022). To achieve a definition that remains meaningful for the non-linear behavior of semi-rigid connections, a characteristic moment (M_d) was introduced, corresponding to 85% of the ultimate moment capacity (M_u) obtained from the tests ($M_d = 0.85 \cdot M_u$) (EN 15512 2022). The rotational stiffness was then evaluated as: $k_0 = 1.15 \cdot M_d / \theta_d$; where θ_d is the rotation measured at the characteristic moment M_d (EN 15512 2022). This procedure enables a consistent and robust quantification of the initial stiffness of the connections prior to the development of pronounced plastic deformations or tab fracture (EN 15512 2022).

2.3. Machine learning methods

To objectively determine the appropriate number of clusters in the PCA-reduced feature space, quantitative clustering validity metrics were employed (Pedregosa et al. 2011). Because unsupervised clustering lacks predefined labels, the quality of a clustering solution cannot be assessed directly and must instead be evaluated through internal consistency measures. The use of multiple clustering metrics provides an objective basis for comparing alternative partitioning schemes and ensures that the selected clustering configuration reflects meaningful structure in the data rather than artefacts of algorithm initialization or model complexity. To determine the op-

timal number of clusters in the PCA-reduced feature space, several clustering validity metrics were employed (Pedregosa et al. 2011; Pilania 2021; Wold et al. 1987). These metrics evaluate complementary aspects of clustering quality and were used jointly to avoid bias associated with reliance on a single criterion. The silhouette score was used to quantify the consistency of cluster assignments by comparing the average intra-cluster distance of each sample with its nearest inter-cluster distance (Shahapure 2020; Shahapure and Nicholas 2020). For a given clustering solution, the silhouette score ranges from −1 to 1, with higher values indicating better-defined and more clearly separated clusters (Pedregosa et al. 2011; Pilania 2021). The elbow method, based on the within-cluster sum of squares (inertia), was applied to assess the reduction in variance achieved by increasing the number of clusters (Shi et al. 2021; Arbelaitz et al. 2013). As inertia decreases monotonically with increasing cluster number, the optimal solution is identified at the point where further increases in cluster count lead to diminishing reductions in inertia. In addition, the Calinski–Harabasz index was computed to evaluate the ratio of between-cluster dispersion to within-cluster dispersion (Calinski and Harabasz 1974; Ikotun et al. 2025). Higher values of this index indicate clustering solutions characterized by compact clusters that are well separated from each other.

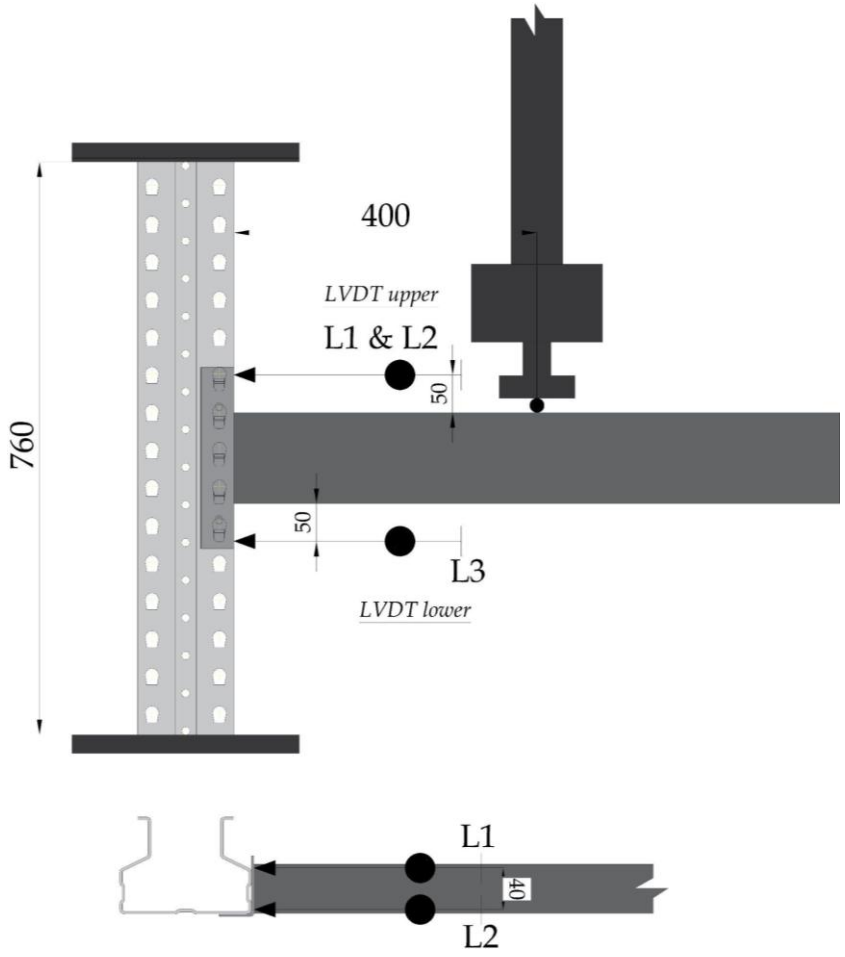

Fig. 2. Schematic test setup showing the boltless beam-to-upright connection, lever arm *a*, gauge length *d*, and LVDT positions.

An initial exploratory analysis was carried out to characterize the statistical structure of the dataset and to support the subsequent development of regression-based machine learning models. The distributions of all input parameters were first examined to identify their range, dispersion, and potential deviations from normality. This step provided insight into the variability of geometric and material features and helped assess the suitability of different regression strategies. To investigate relationships among variables, a Pearson correlation analysis was performed. The resulting correlation matrix quantified the strength and direction of linear dependencies between individual features and the target response. This analysis enabled the identification of parameters most strongly associated with structural capacity, while also revealing limited multicollinearity among inputs. For clarity, redundant elements of the correlation matrix were omitted. In parallel, scatter plots between each input feature and the target variable were generated to visually assess nonlinear trends and detect potential outliers. These visual tools highlighted that several relationships cannot be adequately described by simple linear models, motivating the use of nonlinear regression techniques. Based on these observations, a range of regression models with increasing representational capacity was evaluated. The models considered include linear regressors, regularized linear methods, kernel-based regressors, and ensemble tree-based approaches. Linear

and regularized models provide transparent baselines but are limited in their ability to capture complex interactions. In contrast, ensemble methods such as Random Forest, Extra Trees, Bagging, Gradient Boosting, and XGBoost combine multiple decision trees to model nonlinear behavior and high-order feature interactions more effectively (Yazici and Domínguez-Gutiérrez 2025; Mammadli et al. 2025; Ren et al. 2019; Rahmaty 2023).

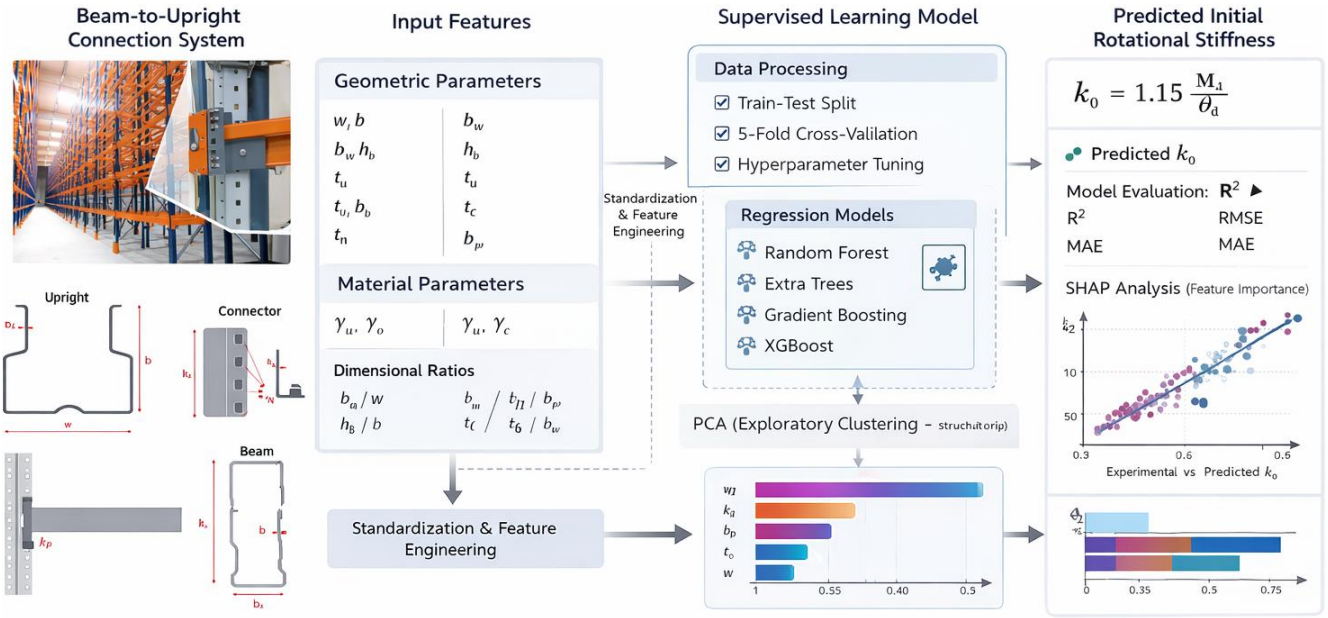

Fig. 3. Schematic of the ML-based rotational stiffness prediction problem for beam-to-upright connections.

Ensemble learning methods improve predictive performance by combining the outputs of multiple base learners, typically decision trees, to reduce model bias, variance, or both. In structural engineering applications, these methods are particularly effective for capturing nonlinear relationships and complex interactions between geometric and mechanical parameters that cannot be represented by single regression models (Mammadli et al. 2025; Ullah et al. 2025). Bagging-based approaches, such as Random Forest and Bagging Regressor, construct multiple decision trees using bootstrap-resampled subsets of the training data. By averaging predictions across independently trained trees, these methods reduce variance and enhance robustness against noise and outliers. Random Forest further introduces random feature selection at each split, which decorrelates individual trees and improves generalization. Boosting-based methods, including Gradient Boosting and XGBoost, adopt a sequential learning strategy in which successive models are trained to correct the residual errors of previous ones (Yazici and Domínguez-Gutiérrez 2025; Ullah et al. 2025). This iterative refinement allows boosting algorithms to focus on difficult-to-predict samples and capture subtle nonlinear trends. Regularization mechanisms and controlled tree depth help mitigate overfitting while maintaining high predictive accuracy. To ensure robust performance assessment, the dataset was divided into training and testing subsets, and model evaluation was further supported by k-fold cross-validation. All models were implemented within a unified machine learning pipeline to ensure consistent preprocessing and fair comparison. This systematic data analysis and visualization stage established both the physical relevance of the selected features and the methodological foundation for reliable regression modelling. To guide the reader through the overall approach, the general workflow of the proposed framework is summarized in Fig. 3.

3. Results

Fig. 4 presents the Pearson correlation matrix of the geometric and mechanical input parameters. The matrix provides a quantitative overview of linear dependencies between variables, with correlation coefficients ranging from −1 (strong negative correlation) to +1 (strong positive correlation). Overall, most parameter pairs exhibit weak to moderate correlations, indicating limited multicollinearity within the dataset. This is advantageous for subsequent multivariate analyses and ML modelling, as it reduces redundancy among predictors. Strong correlations are primarily observed between parameters that are physically related or geometrically coupled. In particular, a very high positive correlation is found between the connector height *hc* and number of tabs *tn* (r≈0.91), reflecting their direct geometric dependence in the connector configuration. Similarly, moderate to strong correlations are observed between *hc* and the base plate parameter *bp* (r≈0.55) and between *tn* and *bp* (r≈0.52). Thickness-related parameters show limited cross-correlation with most geometric dimensions, suggesting that thickness effects introduce additional independent variability. For example, the upright thickness *tu* displays a moderate positive correlation with the width parameter *w* (r≈0.65), while remaining weakly correlated with several other geometric descriptors. Yield strength parameters exhibit moderate correlations among themselves, such as between *yu* and *yb* (r≈0.46) and between *yb* and *yc* (r≈0.45),

consistent with shared material classes across components. Negative correlations are generally weak to moderate and are mainly associated with geometric trade-offs, such as between *bp* and *hb* ($r\approx-0.44$) or between *yu* and *b* ($r\approx-0.45$). These trends indicate compensating design choices rather than strict dependencies.

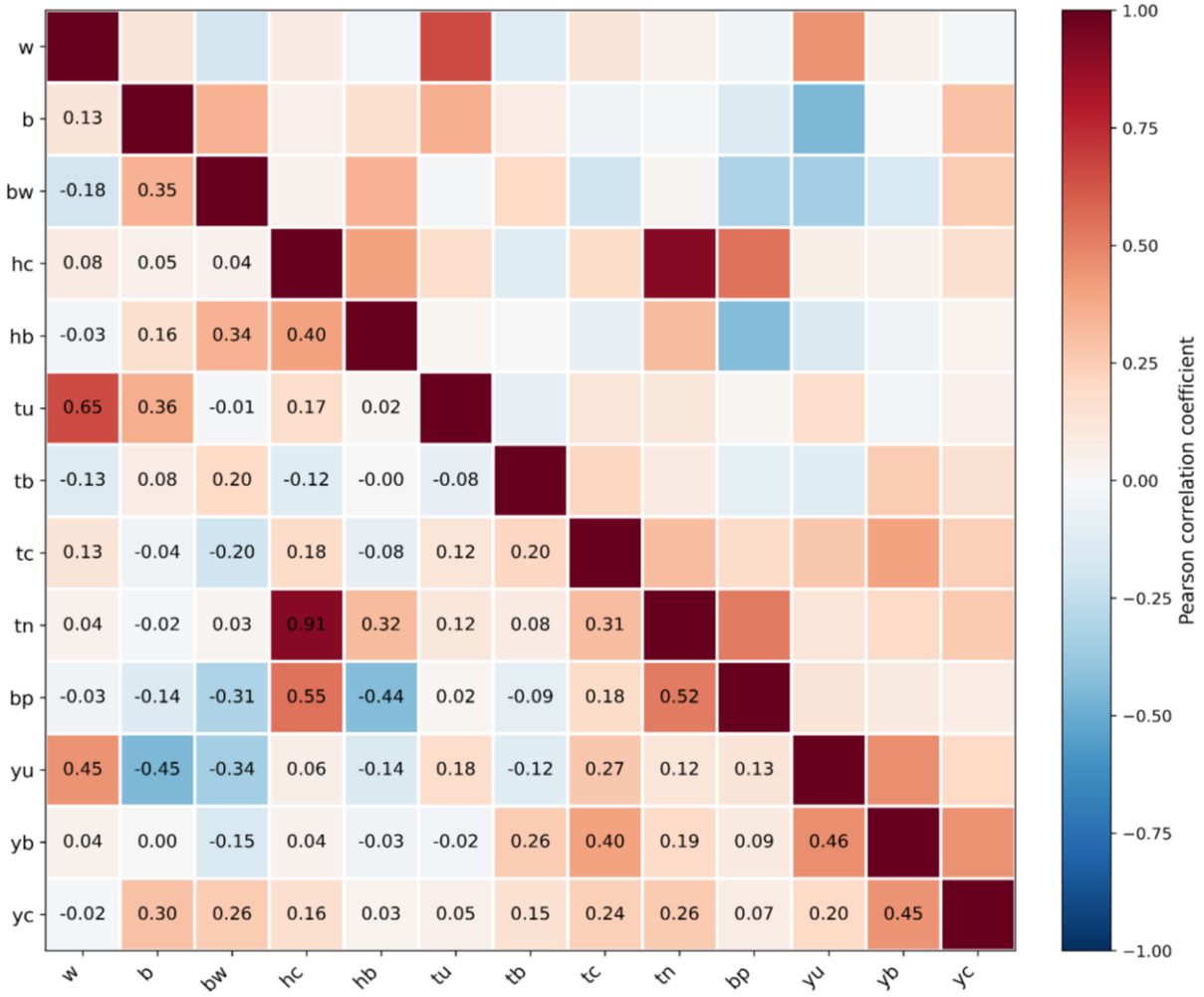

Fig. 4. Pearson correlation matrix of the geometric and mechanical input parameters.

Although a strong correlation (r=0.91) is observed between connector height (*hc*) and number of tabs (*tn*), this reflects inherent geometric coupling in connector design. Tree-based ensemble regression models are relatively robust to multicollinearity, as they do not rely on global coefficient estimation. Nevertheless, correlated features may share importance in SHAP analysis, and this redistribution should be interpreted as reflecting shared geometric influence rather than independent physical effects.

3.1. Classification

To ensure a robust and defensible identification of cluster structure, multiple clustering validity metrics were employed rather than relying on a single criterion. The silhouette score was used to evaluate the balance between intra-cluster cohesion and inter-cluster separation, providing a direct measure of how well individual samples are assigned to their respective clusters (Arbelaitz et al. 2013). The elbow criterion, based on the within-cluster sum of squares (inertia), was adopted to assess the reduction in variance achieved by increasing the number of clusters and to identify diminishing returns in model complexity. In addition, the Calinski–Harabasz index was considered to quantify the ratio of between-cluster dispersion to within-cluster dispersion, thereby favouring solutions that maximize separation while maintaining compact clusters (Pedregosa et al. 2011). The combined use of these complementary metrics mitigates the limitations inherent to any single measure and enhances the reliability of cluster selection.

Fig. 5 compares the normalized values of the silhouette score, inverted inertia (elbow method), and Calinski–Harabasz index as a function of the number of clusters. All three metrics exhibit a generally increasing trend with increasing k, reflecting the expected improvement in cluster compactness and separation as additional clusters are introduced. However, the rate of improvement differs among the metrics. The silhouette score shows a pronounced increase up to approximately k=6, beyond which further gains become more gradual. A similar behaviour is observed for the inverted inertia, indicating that variance reduction becomes less significant for larger numbers of clusters. The Calinski–

Harabasz index follows the same overall trend but displays a more conservative increase at lower k, highlighting its stronger sensitivity to over-partitioning. The convergence of the three metrics at higher k values suggests improved clustering quality but also underscores the need to balance statistical improvement against physical interpretability, motivating the subsequent analysis of cluster structure in the reduced PCA space.

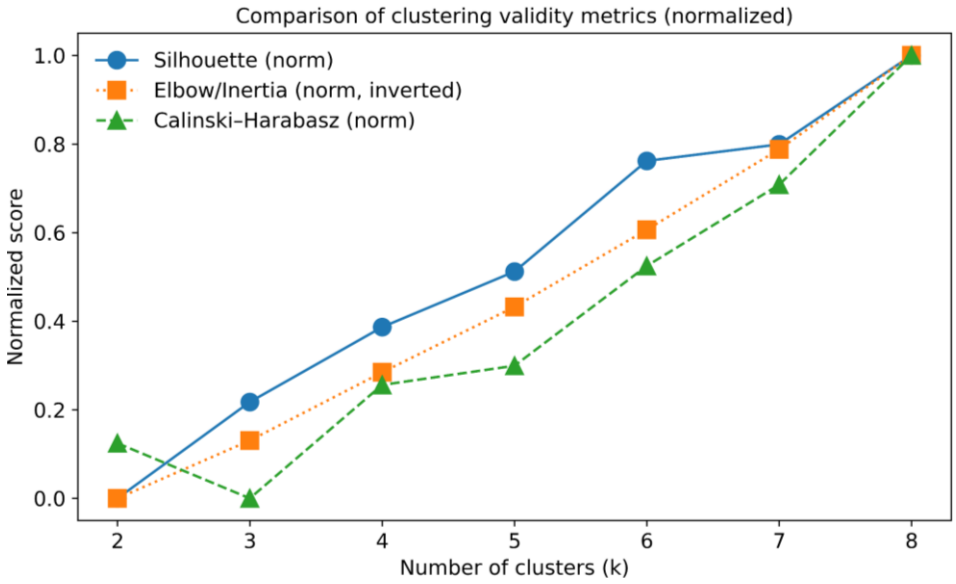

Fig. 5. Normalized comparison of silhouette score, inverted inertia (elbow method), and Calinski–Harabasz index used to assess clustering validity as a function of the number of clusters.

Fig. 6 presents the clustering results obtained by applying K-means clustering in the PCA-reduced feature space. The first two principal components, which explain 22.1 % and 17.4 % of the total variance and reflect the multivariate nature of the governing geometric and mechanical parameters, are shown in all panels to facilitate direct comparison between clustering strategies. Panels (a)–(c) display the PCA score plots colored according to cluster membership, where the number of clusters was selected using different clustering validity criteria. In panel (a), the silhouette-based selection (k=8) results in a relatively fine partitioning of the dataset, with several compact clusters distributed across the PCA space. Panel (b) shows the clustering obtained using the elbow criterion (k=4), which yields a coarser segmentation and merges several groups that are separated in the silhouette-based solution. Panel (c), corresponding to the Calinski–Harabasz criterion (k=8), produces a cluster structure similar in granularity to the silhouette-based case, with comparable spatial separation of clusters, although individual cluster assignments differ locally. Panel (d) shows the same PCA score plot colored by the continuous stiffness parameter k_0. A clear non-uniform distribution of k_0 values is observed across the PCA space, with regions of higher and lower stiffness occupying distinct areas. This representation highlights systematic variations of k_0 across the dataset and provides a direct visual link between the multivariate structure captured by the principal components and the mechanical response of the connections.

The PCA–K-means clustering results reveal that the dataset is naturally structured into groups that reflect distinct mechanical response regimes of the beam-to-upright connections. The finer clustering obtained using the silhouette- and Calinski–Harabasz-based selections (k=8) suggests that multiple mechanically distinct regimes coexist within the dataset. These clusters likely correspond to variations in connector geometry and thickness combinations that control local bending, bearing, and engagement behaviour between tabs, beam flanges, and upright perforations. In particular, clusters that are well separated along PC1, which captures the largest fraction of variance, are associated with changes in global geometric proportions and connector dimensions, indicating a transition between more compliant and stiffer joint configurations. Separation along PC2 reflects secondary effects, such as thickness ratios and material strength variations, which modulate stiffness within otherwise similar geometric layouts. In contrast, the elbow-based solution (k=4) produces a coarser partitioning that groups together configurations exhibiting comparable overall stiffness behaviour but potentially different local deformation patterns. From a mechanical standpoint, this indicates that while several configurations may share similar global stiffness levels, their underlying load transfer mechanisms, such as the balance between connector bending and local bearing deformation, can differ. The finer clustering therefore captures subtler mechanical distinctions that are not resolved when only global variance reduction is considered.

The PCA score plot colored by the stiffness parameter k_0 further supports this interpretation. Regions of high and low k_0 values are not randomly distributed but instead form continuous gradients across the PCA space, demonstrating that stiffness evolves systematically with changes in geometry and material properties. High-stiffness regions correspond to clusters characterized by increased effective load-bearing area, higher thickness values, and stronger material combinations, which reduce local compliance and delay the onset of connector defor-

mation. Conversely, lower k_0 regions are associated with more slender geometries or reduced connector engagement, leading to increased rotational flexibility. Overall, the clustering analysis highlights that the mechanical behavior of pallet rack connections cannot be described by a single stiffness class but instead spans multiple regimes governed by interacting geometric and material parameters. The identification of these regimes provides a physically meaningful framework for interpreting the data-driven model results and supports the use of machine-learning approaches to capture the nonlinear and multivariate nature of semi-rigid joint behavior. It should be noted that the PCA–K-means clustering was conducted for exploratory analysis and structural interpretation of the dataset. The regression models presented in the next section are trained using the original physically meaningful geometric and mechanical parameters rather than the PCA-transformed components, in order to preserve interpretability and avoid loss of mechanical transparency.

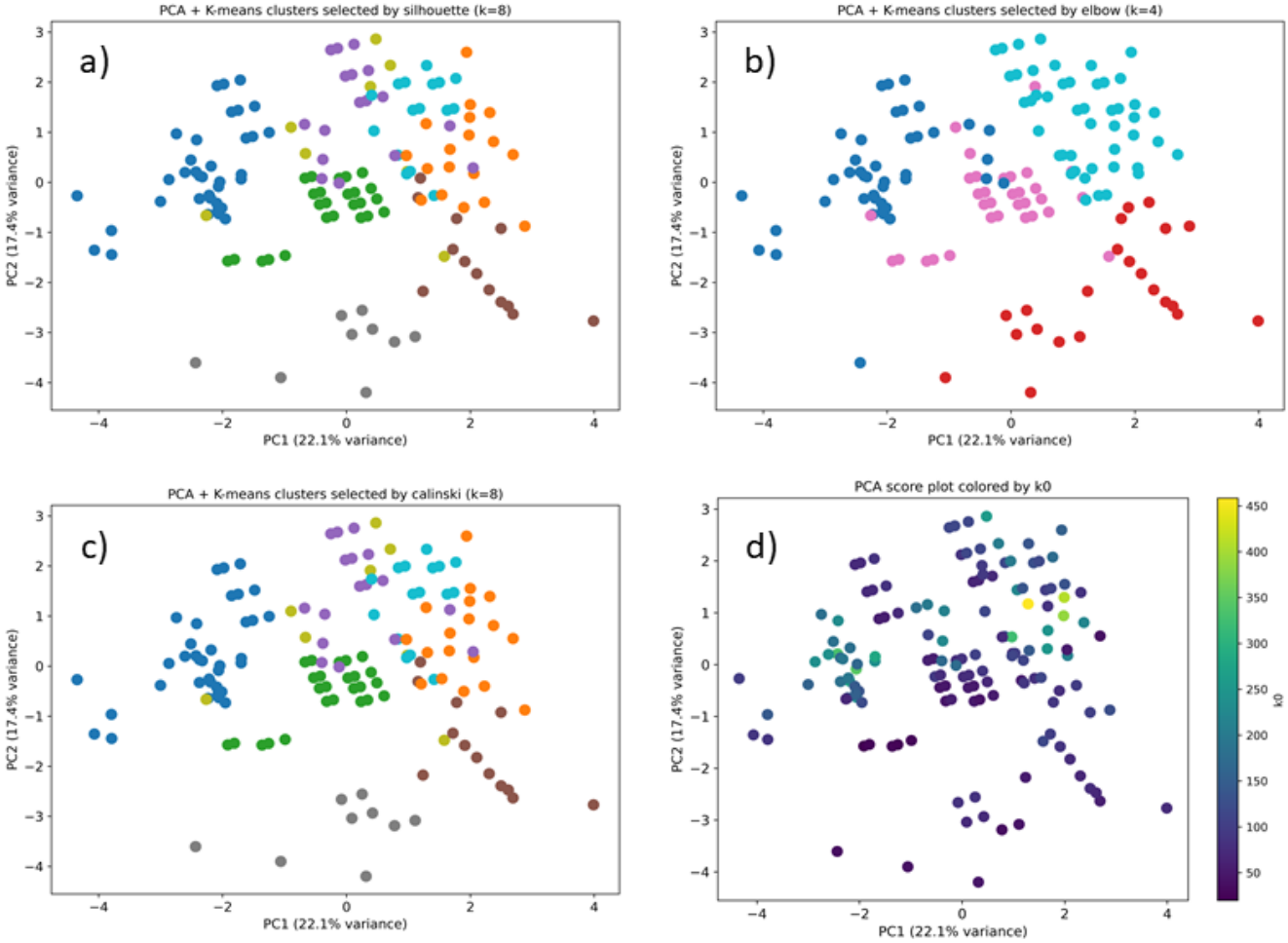

Fig. 6. PCA score plots showing K-means clustering of beam-to-upright connection configurations using different cluster selection criteria: (a) silhouette method (k=8); (b) elbow method (k=4); (c) Calinski–Harabasz index (k=8); and (d) PCA scores colored by the initial rotational stiffness k_0.

3.2. Regression

Fig. 7 compares the predictive performance of the evaluated regression models using three complementary metrics: the coefficient of determination (R^2), the root mean square error (RMSE), and the mean absolute error (MAE). The results demonstrate clear differences in accuracy between linear, nonlinear, and ensemble-based approaches. Linear and regularized linear models (Linear, Ridge, Lasso, and Elastic Net) exhibit moderate predictive capability, with (R^2) values clustered around 0.63–0.68 and relatively high RMSE and MAE values. This indicates that purely linear relationships are insufficient to fully capture the dependence of the initial rotational stiffness on the combined geometric and material input parameters. Bayesian Ridge and PLS regression further show reduced performance, with lower (R^2) values and higher error metrics, reflecting limited flexibility in representing complex interactions. Tree-based models provide a marked improvement in predictive accuracy. The single Decision Tree achieves a substantial increase in ($R^2 \approx 0.83$) and a corresponding reduction in both RMSE and MAE, highlighting the importance of nonlinear feature partitioning. However, ensemble tree methods consistently outperform individual models. Random Forest, Bagging, Gradient Boosting, XGBoost, and Extra Trees all achieve high (R^2) values exceeding 0.90, accompanied by significantly reduced error magnitudes. Among these, Extra Trees yields the best overall performance, with the highest (R^2) and the lowest RMSE and MAE, closely followed by Gradient Boosting, XGBoost, and Bagging.

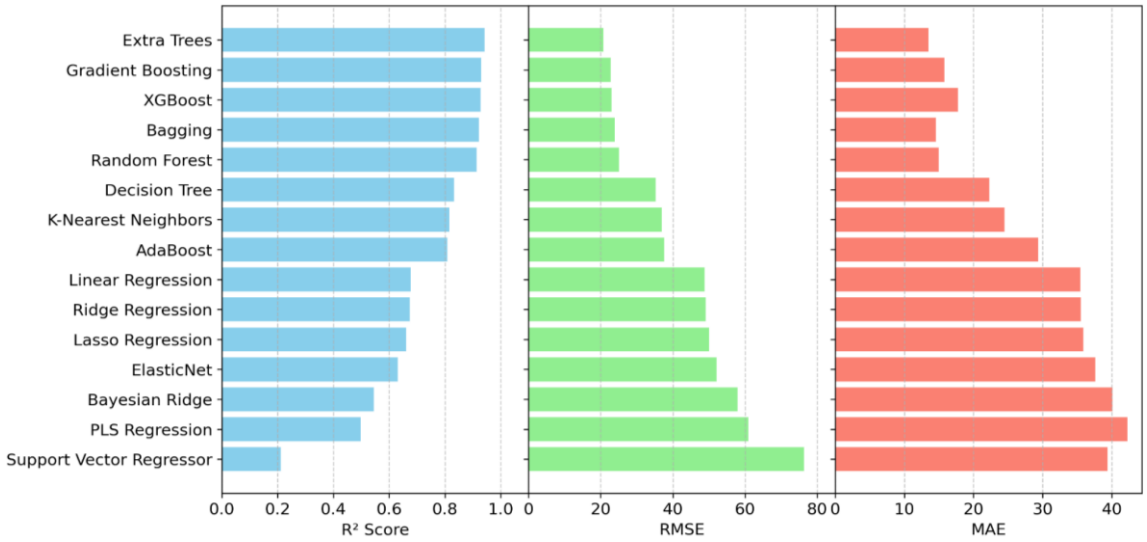

Fig. 7. Comparison of regression model performance in predicting the initial rotational stiffness k_0, evaluated using R^2, RMSE, and MAE.

These results indicate that ensemble averaging and boosting strategies are particularly effective in capturing the multivariate and nonlinear relationships governing connection stiffness. In contrast, the Support Vector Regressor exhibits poor performance across all metrics, suggesting inadequate model tuning or limited suitability for the present dataset. For reference, the data of the performance of each model is presented in Tab 1. Overall, the figure demonstrates that ensemble-based regression models provide a substantial accuracy advantage over linear and single-model approaches for predicting the initial rotational stiffness, motivating their selection for subsequent analysis and model interpretation.

Table 1. Performance comparison of regression models for prediction of the initial rotational stiffness (k_0).

Model	RMSE	MAE	(R^2)
Linear Regression	48.82	35.45	0.68
Ridge Regression	49.07	35.52	0.67
Lasso Regression	50.06	35.87	0.66
Elastic Net	52.18	37.62	0.63
Bayesian Ridge	57.98	40.10	0.55
Decision Tree	35.17	22.25	0.83
Random Forest	25.12	14.99	0.91
Extra Trees	20.77	13.49	0.94
Gradient Boosting	22.74	15.78	0.93
AdaBoost	37.59	29.35	0.81
Bagging	23.93	14.57	0.92
K-Nearest Neighbors	36.94	24.46	0.82
Support Vector Regressor	76.38	39.36	0.21
XGBoost	23.05	17.74	0.93
PLS Regression	60.96	42.23	0.50
Model	RMSE	MAE	(R2)

Fig. 8 compares the predicted and experimentally measured values of the initial rotational stiffness k_0 for the best-performing ensemble regression models, namely Random Forest, Extra Trees, Gradient Boosting, Bagging, and XGBoost. Each point represents a single test sample, while the dashed red line indicates the ideal one-to-one correspondence between predicted and actual values. Across the full stiffness range, the majority of predictions cluster closely around the perfect-fit line, demonstrating a strong agreement between model outputs and reference values. This trend is particularly pronounced in the low to intermediate stiffness regime, where all ensemble models show limited scatter and small deviations from the ideal prediction. In this region, the predicted values follow the linear trend with high consistency, indicating reliable model performance for the most frequently represented configurations in the dataset. At higher stiffness values, the dispersion of predictions increases moderately, reflecting the reduced density of samples in this range. Nevertheless, the ensemble models maintain good predictive accuracy, with most points remaining near the one-to-one line and no systematic bias toward overestimation or underestimation being observed. Minor deviations between individual models are visible, but their overall predictive behavior remains comparable. It can be observed in Fig. 8 that the dispersion between predicted and measured values increases in the higher stiffness range. This behavior is primarily attributed to the lower density of experimental data in this regime, as most configurations in the compiled database correspond to low- to intermediate-stiffness connections. The relative scarcity of high-stiffness samples limits the model's exposure to these configurations during training and increases prediction uncertainty. In addition, high-stiffness joints are typically associated with thicker sections and more robust connector geometries, where the dominant deformation mechanisms may transition from tab bending–controlled behavior to more localized bearing or upright distortion effects. Such shifts can introduce additional nonlinear interactions between geometric and material pa-

rameters, increasing variability. Future expansion of the database, particularly with additional high-stiffness experimental configurations, would further enhance predictive robustness in this range. In addition, model performance was evaluated using 5-fold cross-validation with random shuffling of samples prior to partitioning. For the dataset size (n=151), k=5 provides a suitable balance between training data sufficiency and validation robustness. The database contains no duplicate specimens; however, future work may consider grouped cross-validation based on study source or manufacturer to further assess cross-family generalization capability.

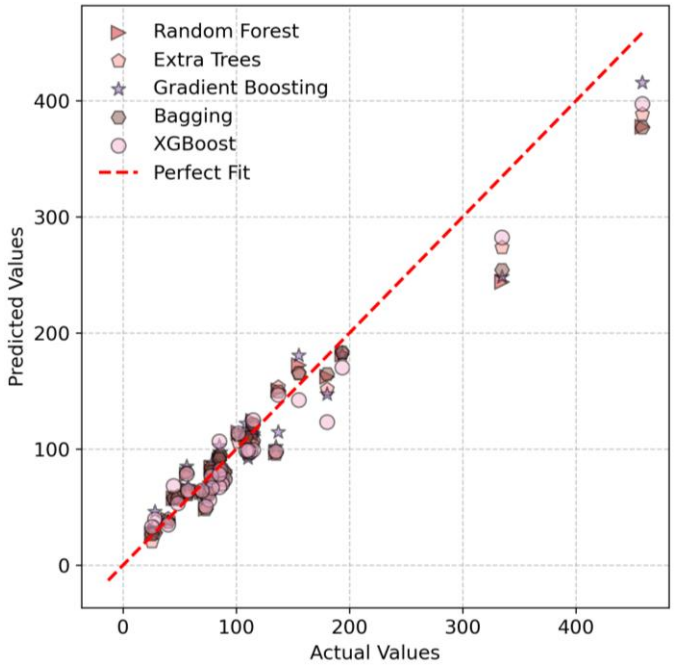

Fig. 8. Comparison of predicted and measured initial rotational stiffness k_0 for the selected ensemble regression models, with the dashed line indicating perfect agreement.

Fig. 9(a) presents the SHAP summary plot for the selected ensemble regression model, illustrating the relative importance of the input features and their contribution to the predicted initial rotational stiffness k_0. Features are ranked from top to bottom according to their mean absolute SHAP value, indicating their overall influence on the model output. Each point represents an individual sample, with the horizontal position corresponding to the SHAP value (positive or negative impact on k_0) and the color indicating the magnitude of the corresponding feature value. The results show that geometric parameters associated with the beam and connector configuration dominate the prediction of k_0. In particular, the beam height *hb* exhibits the highest influence, with a wide spread of SHAP values spanning both positive and negative contributions. Higher values of *hb* are predominantly associated with positive SHAP values, indicating an increase in predicted stiffness, while lower values tend to reduce k_0. Yield strength of the beam (*yb*) and connector thickness *tc* also show strong contributions, with clear separation between low- and high-value regimes and consistent directional effects on the model output. Thickness-related parameters, such as the upright thickness *tu* and tab or notch thickness *tn*, contribute moderately to the prediction, generally exhibiting positive SHAP values at higher feature magnitudes. In contrast, parameters such as base plate width bp, beam thickness *tb*, and connector height *hc* show more localized effects, with narrower SHAP distributions centred closer to zero, suggesting secondary influence compared to the dominant geometric variables. Material strength parameters display a mixed but generally smaller impact. While the connector and upright yield strengths (*yc* and *yu*) influence the predictions, their SHAP value distributions are more compact, indicating that variations in stiffness are less sensitive to strength changes than to geometric modifications within the studied range.

Fig. 9(b) presents the global feature importance derived from the SHAP analysis of the Extra Trees regression model. The bars represent the mean absolute SHAP value for each input parameter, quantifying its average contribution to the predicted initial rotational stiffness k_0 across the dataset. Features are ranked in descending order of importance, providing a clear hierarchy of the parameters influencing the model output. The results indicate that the beam height *hb* is the most influential parameter, exhibiting a substantially higher mean SHAP value than all other features. This highlights the dominant role of beam geometry in controlling the rotational stiffness of the beam-to-upright connection. The beam yield strength *yb* and the connector thickness *tc* follow as the next most important contributors, emphasizing the combined influence of material resistance and local connector stiffness on the joint response. Upright thickness *tu* and connector yield strength *yc* also show significant contributions, indicating that both structural stiffness and material strength at the connection interface play a relevant role in determining k_0. Geometric parameters such as the beam width *w* and notch or tab thickness *tn* exhibit moderate importance, suggesting a sec-

ondary but non-negligible influence on rotational behavior. In contrast, parameters including base plate width *bp*, beam thickness *tb*, connector height *hc*, upright yield strength *yu*, and beam width parameter *b* contribute less to the prediction, while the beam web width *bw* shows the lowest impact among the considered inputs. These lower-ranked features primarily modulate the stiffness response rather than governing it directly within the studied parameter range. Although yield strength appears less influential than geometric parameters in the SHAP ranking, this should not be interpreted as mechanical irrelevance. In thin-walled cold-formed members, strength and local buckling behaviour are intrinsically coupled through plate slenderness and post-yield response. The lower statistical importance observed here reflects the relative variability within the dataset rather than a fundamental decoupling of strength from stiffness mechanisms.

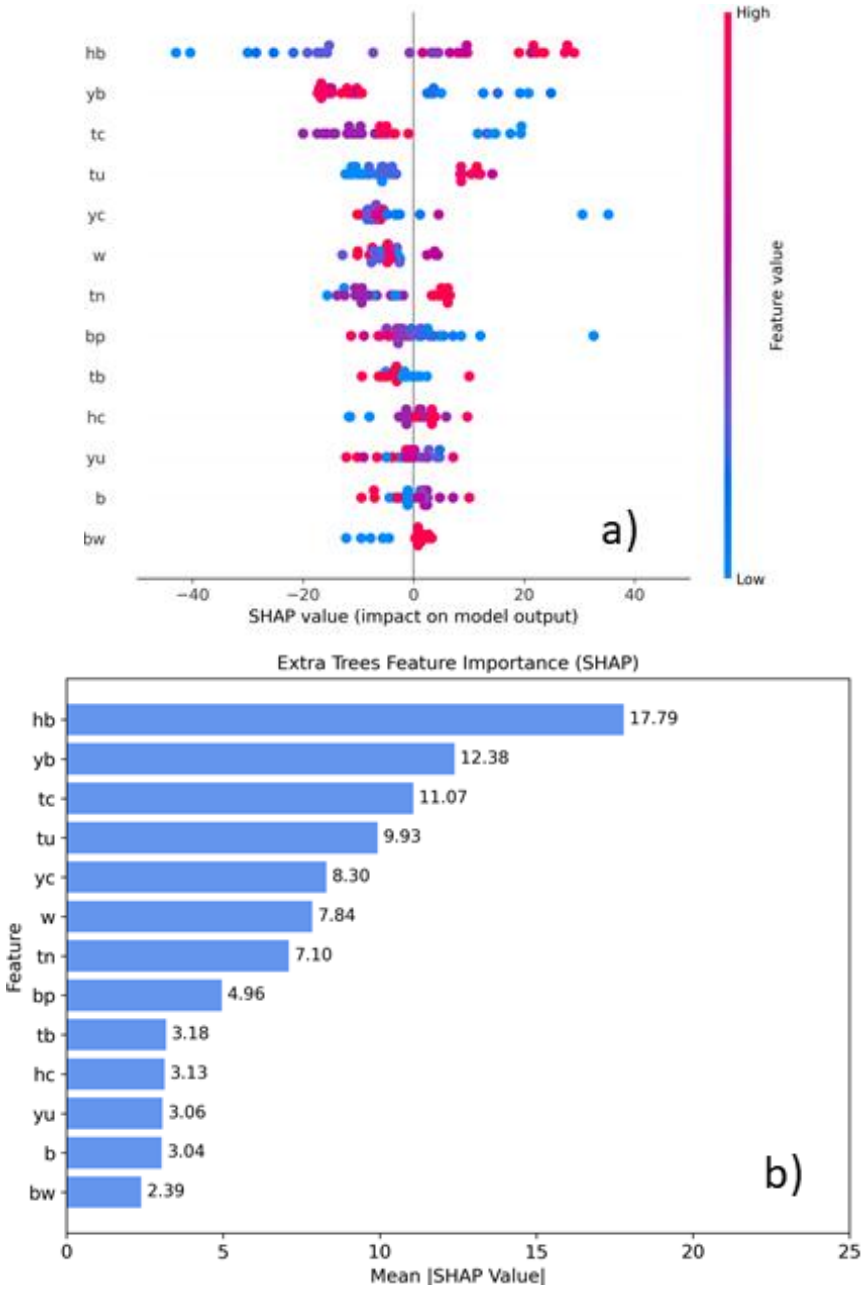

Fig. 9. SHAP summary plot showing the relative importance and directional influence of input features on the predicted initial rotational stiffness k_0 for SHAP values in (a) and Mean in (b).

4. Conclusions

This study demonstrates the effectiveness of data-driven machine-learning approaches for predicting the initial rotational stiffness k_0 of beam-to-upright connections in steel pallet rack systems using only geometric and mechanical input parameters. By combining experimental data with modern regression techniques, the proposed framework provides an accurate and efficient alternative to traditional analytical formulations and costly laboratory testing. A systematic comparison of linear, nonlinear, and ensemble-based regression models revealed that ensemble tree methods significantly outperform linear and single-model approaches. Among the

evaluated models, Extra Trees, Gradient Boosting, Random Forest, and XGBoost achieved the highest predictive accuracy, with coefficients of determination exceeding 0.9 and substantially reduced prediction errors. These results highlight the strong nonlinear character of the relationship between connection stiffness and its governing parameters, which cannot be adequately captured by simplified analytical or linear models. Unsupervised learning techniques further supported the analysis by revealing intrinsic structure within the dataset. PCA-based clustering identified distinct regimes in the reduced feature space, corresponding to different combinations of geometric proportions and material properties that govern joint behaviour. The use of multiple clustering validity metrics ensured robust and physically consistent cluster selection. The application of SHAP analysis provided model transparency and physical interpretability, demonstrating that the machine-learning models rely primarily on geometrical parameters controlling load transfer and deformation mechanisms, such as beam height, connector thickness, and upright dimensions, while material strength parameters play a secondary but complementary role. This consistency with established mechanical understanding reinforces the reliability of the learned relationships. Overall, the results highlight the key advantages of machine-learning methods for the analysis of semi-rigid structural connections: the ability to capture complex multivariate and nonlinear interactions, to generalize across a wide range of configurations, and to provide accurate predictions without the need for extensive experimental testing. The proposed framework enables more realistic numerical modelling, supports the adoption of semi-rigid design concepts, and offers a practical pathway toward automated and performance-based design of steel storage systems.

Acknowledgements

None declared.

Funding

This research was funded through the European Union Horizon 2020 research and innovation program under Grant Agreement No. 857470 and from the European Regional Development Fund under the program of the Foundation for Polish Science International Research Agenda PLUS, Grant No. MAB PLUS/2018/8, and the initiative of the Ministry of Science and Higher Education "Support for the activities of Centers of Excellence established in Poland under the Horizon 2020 program" under Agreement No. MEiN/2023/DIR/3795.

Conflict of Interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this manuscript.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available but are available from the corresponding author upon reasonable request.

AI Assistance

No AI-based tools were used in the preparation of this manuscript.

Author Contributions

All authors made substantial contributions to the conception and design of the study, acquisition of data, analysis and interpretation of data; drafted or critically revised the manuscript for important intellectual content; and approved the final version to be published.

REFERENCES

Arbelaitz O, Gurrutxaga I, Muguerza J, Pérez JM, Perona I (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46(1), 243–256.

Calinski T, Harabasz J (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1–27.

Dai L, Zhao X, Rasmussen KJR (2018). Cyclic performance of steel storage rack beam-to-upright bolted connections. *Journal of Constructional Steel Research*, 148, 28–48.

EN 15512 (2022). Steel static storage systems Adjustable pallet racking systems Principles for structural design. European Committee for Standardization (CEN), Brussels.

Ikotun AM, Habyarimana F, Ezugwu AE (2025). Benchmarking validity indices for evolutionary K-means clustering performance. *Scientific Reports*, 15, 21842.

Mammadli B, Yazici C, Gürbüz M, Kocaman İ, Dominguez-Gutierrez FJ, Özkal FM (2025). A data-driven machine learning approach for predicting axial load capacity in steel storage rack columns. *Results in Engineering*, 28, 107475.

Mohan V, Prabha P, Rajasankar J, Iyer NR, Raviswaran N, Nagendiran V, Kamalakannan SS (2015). Cold-formed steel pallet rack connection: An experimental study. *International Journal of Advanced Structural Engineering*, 7(1), 55–68.

Mohan V, Vishnu CR (2013). Joint stiffness of cold-formed steel pallet rack connections: A comparison of the methodology. *Journal of Structural Engineering (Madras)*, 40(5), 457–465.

Özkal FM, Yazici C (2024). Effects of steel bolts on retrofitting the buckling performance of storage rack system uprights. *Structures*, 64, 106578.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Pilania G (2021). Machine learning in materials science: From explainable predictions to autonomous design. *Computational Materials Science*, 193, 110360.

Prabha P, Marimuthu V, Saravanan M, Jayachandran SA (2010). Evaluation of connection flexibility in cold formed steel racks. *Journal of Constructional Steel Research*, 66(7), 863–872.

Rahmaty M (2023). Machine learning with big data to solve real-world problems. *Journal of Data Analysis*, 2(1), 9–16.

Ren J, Chen X, Tan Y, Liu D, Duan M, Liang L, Qiao L (2019). Archivist: A machine learning assisted data placement mechanism for hybrid storage systems. In: Proceedings of the 2019 IEEE International Conference on Computer Design (ICCD), 676–679.

Rosli NAM, Tan CG, Hashim H, Sulong NHR (2025). Experimental investigation and parameter analysis of boltless steel pallet rack beam-to-column connection. *Journal of Constructional Steel Research*, 231, 109578.

Santamaria D, Oyanguren A, Iñurritegui A, Larrañaga J, Ulacia I (2024). Numerical characterisation of reaction forces' contribution to initial stiffness in speed-lock beam-to-upright connections. *Journal of Constructional Steel Research*, 223, 108995.

Shah SN, Sulong NH, Shariati M, Jumaat MZ (2015). Steel rack connections: Identification of most influential factors and a comparison of stiffness design methods. *PLoS One*, 10(10), e0139422.

Shah SNR, Ramli Sulong NH, Khan R, Jumaat MZ (2017). Structural performance of boltless beam end connectors. *Advanced Steel Construction*, 13(2), 144–159.

Shah SNR, Sulong NHR, Khan R, Jumaat MZ, Shariati M (2016). Behavior of industrial steel rack connections. *Mechanical Systems and Signal Processing*, 70–71, 725–740.

Shahapure KR (2020). Cluster quality analysis using silhouette score. M.S. Writing Project, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County (UMBC).

Shahapure KR, Nicholas C (2020). Cluster quality analysis using silhouette score. In: Proceedings of the IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), Sydney, NSW, Australia, 747–748.

Shariati M, Tahir MM, Wee TC, Shah SNR, Jalali A, Abdullahi MM, Khorami M (2018). Experimental investigations on monotonic and cyclic behavior of steel pallet rack connections. *Engineering Failure Analysis*, 85, 149–166.

Shi C, Wei B, Wei S, Wang W, Liu H, Liu J (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *Journal of Wireless Communications and Networking*, 2021, 31.

Ślęczka L, Kozłowski A (2007). Experimental and theoretical investigations of pallet racks connections. *Advanced Steel Construction*, 3(2), 607–627.

Ullah N, Guzmán-Aroca F, Martínez-Álvarez F, De Falco I, Sannino G (2025). A novel explainable AI framework for medical image classification integrating statistical, visual, and rule-based methods. *Medical Image Analysis*, Article 103665.

Wold S, Esbensen K, Geladi P (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1–3), 37–52.

Yazici C, Arik E, Özkal FM (2023). Influence of design parameters on beam-to-column connections of steel storage rack systems. *Engineering Failure Analysis*, 152, 107439.

Yazici C, Domínguez-Gutiérrez FJ (2025). Machine learning techniques for estimating high–temperature mechanical behavior of high strength steels. *Results in Engineering*, 25, 104242.

Yazici C, Gusella F, Özkal FM (2026). Effect of beam-to-connector weld position on the flexural response of boltless rack connections. *Journal of Constructional Steel Research*, 237(Part B), 110141.

Yin L, Tang G, Zhang M, Wang B, Feng B (2016). Monotonic and cyclic response of speed-lock connections with bolts in storage racks. *Engineering Structures*, 116, 40–55.

Zhao X, Dai L, Rasmussen KJR (2018). Hysteretic behaviour of steel storage rack beam-to-upright boltless connections. *Journal of Constructional Steel Research*, 144, 81–105.

Zhao X, Wang T, Chen Y, Sivakumaran KS (2014). Flexural behavior of steel storage rack beam-to-upright connections. *Journal of Constructional Steel Research*, 99, 161–175.