Engineering with Computers (2023) 39:3671–3689
https://doi.org/10.1007/s00366-022-01781-9

ORIGINAL ARTICLE

A novel outlier-insensitive local support vector machine for robust data-driven forecasting in engineering

Huan Luo[1] · **Stephanie German Paal**[2]

Received: 23 January 2022 / Accepted: 30 December 2022 / Published online: 6 January 2023

Abstract
Machine learning (ML)-based data-driven methods have promoted the progress of modeling in many engineering domains. These methods can achieve high prediction and generalization performance for large, high-quality datasets. However, ML methods can yield biased predictions if the observed data (i.e., response variable y) are corrupted by outliers. This paper addresses this problem with a novel, robust ML approach that is formulated as an optimization problem by coupling locally weighted least-squares support vector machines for regression (LWLS-SVMR) with one weight function. The weight is a function of residuals and allows for iteration within the proposed approach, significantly reducing the negative interference of outliers. A new efficient hybrid algorithm is developed to solve the optimization problem. The proposed approach is assessed and validated by comparison with relevant ML approaches on both one-dimensional simulated datasets corrupted by various outliers and multi-dimensional real-world engineering datasets, including datasets used for predicting the lateral strength of reinforced concrete (RC) columns, the fuel consumption of automobiles, the rising time of a servomechanism, and dielectric breakdown strength. Finally, the proposed method is applied to produce a data-driven solver for computational mechanics with a nonlinear material dataset corrupted by outliers. The results all show that the proposed method is robust against non-extreme and extreme outliers and improves the predictive performance necessary to solve various engineering problems.

Keywords Data-driven methods · Support vector machines · Robust regression · Outliers · Locally weighted least squares · Nonlinear elasticity

1 Introduction

Data-driven methods, where progress in an activity (e.g., prediction or decision-making) is driven by data instead of being driven by personal experience and inference, have attracted considerable interest and also achieved great success in the engineering fields [1–4]. For example, many such methods [e.g., machine learning (ML) techniques] have been applied to: extract constitutive manifolds [5–10], identify the stress–strain relation [11–15], produce data-driven solvers [16–19], and predict the strength and deformation of engineering structures and materials [20–23]. Typically, the existing ML-based data-driven approaches are able to fit and generalize the input data well and can produce extremely good prediction capabilities if the input data are high quality and reasonably large in size [24]. However, if the input data are corrupted by outliers, these data-driven methods (e.g., methods for regression), especially those that are sensitive to outliers, will yield unreliable prediction. Some of them even break down when the data are contaminated by extreme outliers. Outliers are those observations that are far away from all other observations due to misplaced decimal points, recording or transmission errors, or exceptional phenomena. These are all common occurrences in real-world data [25]. In general, there are two commonly employed ways to deal with the outliers for regression problems [25]. The first is to use robust regression approaches, while the second method is to construct outlier diagnostics. A robust regression approach first fits a regression model that adequately addresses the normal data points and then discovers the outliers as those points having large residuals estimated from

✉ Huan Luo
hluo@ctgu.edu.cn

Stephanie German Paal
spaal@civil.tamu.edu

[1] College of Civil Engineering and Architecture, China Three Gorges University, Yichang 443002, Hubei, China

[2] Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77843, USA

the robust regression model [26–28]. On the contrary, outlier diagnostics first identify the outliers and then remove them and fit the remaining normal data points [29–32]. In some applications, both methods yield exactly the same result. However, outlier diagnostics may result in outliers which are not entirely detected, leading to biased results, while robust regression does not pose such a risk. For this reason, attention of this paper is only focused on the robust regression approaches.

Robust regression approaches include least absolute deviations (LAD), least trimmed squares (LTS), *M*-estimators, etc., which were proposed to address the fact that the least-squares (LS) method is easily affected by outliers [25]. These robust methods were originally developed for parametric regression (e.g., linear regression). Recently, many efforts have been made to incorporate these regression approaches into the reformulation of data-driven methods to enhance their robustness. For example, LTS has been integrated into backpropagation neural networks (BPNNs) to replace the mean squared error (MSE) as the minimization criterion [33], and LAD has been applied in random forests (RFs) to replace the original LS as the splitting rule for building the regression trees [34]. Least-squares support vector machines for regression (LS-SVMR) [35] is one of the more frequently used data-driven regression methods in civil engineering disciplines [36–47]. The LS-SVMR [35] is a reformulation of SVMR [48], which uses the sum of squared errors (SSE) as the loss function and equality constraints in place of inequality constraints to greatly simplify the SVMR formulation. Due to this, the LS-SVMR solves a linear system problem instead of the complex quadratic programming (QP) problem, leading to greater computational efficiency. However, the use of SSE as the loss function in the formulation of LS-SVMR leads to a non-robust property. To overcome this problem, the weighted SSE has been adapted as the loss function by Suykens et al. [49] to substitute the original SSE for the reformulation of LS-SVMR, which resulted in the new LS-SVMR variant called WLS-SVMR that is robust to outliers. The weight function used in WLS-SVMR is a function of residuals estimated by LS-SVMR, where the potential outliers tend to have larger residuals. The points which have larger residuals in the training set will then be assigned smaller weights to reduce their associated negative influence. However, WLS-SVMR breaks down under non-Gaussian noise distribution with heavy tails (i.e., extreme outliers) [50]. To solve this problem, De Brabanter et al. [50] proposed an iterative version of WLS-SVMR (IWLS-SVMR), where the weights are updated in each iteration to reduce the negative influence of extreme outliers until convergence criteria are reached.

Both WLS-SVMR and IWLS-SVMR are robust, *global* data-driven regression models, meaning that their solution requires the fitting of the entire training set. However, in many cases, the performance of global models can be further improved by local models [51–55]. As introduced in Bottou and Vapnik [53], the local learning algorithms attempt to locally adjust the capacity of the training system to the properties of the training set in each area of the input space. This results in a local model that only requires the fitting of a subset of the training data nearby (relevant to) the query point and can overcome the potential negative influence of irrelevant points. Therefore, a robust, *local* model may provide an improvement under these circumstances when compared to the robust, *global* models. This is because the robust, local model can yield a model that both overcomes the negative interference of outliers and avoids the potential negative influence of irrelevant points, achieving a suitable trade-off between the capacity of the learning system and the number of training data points [46, 53–55]. A local version of LS-SVMR, called moving LS-SVMR (M-LS-SVMR), was proposed by Karevan et al. [55] for weather temperature prediction. In this method, the Gaussian- and cosine-based weight functions are used to measure the similarity between training sample and test data. However, this method does not use a robust weighting scheme to reduce the negative influence of outliers. Further, the majority of existing local models are based on polynomial regression, which means that the local models are polynomial functions [56–59]. These local models are not real data-driven regression methods because of the assumption of polynomial functions within local regions. Thus, the families of such local models are out of the scope of this paper. One local data-driven regression model, called locally weighted LS-SVMR (LWLS-SVMR), was recently developed for generalized prediction of the drift capacity of RC columns [46]. LWLS-SVMR [46] integrates LS-SVMR with a locally weighted learning algorithm [56–59] to locally adjust the capacity of LS-SVMR to the properties of the training set in each area of the input space, thus enhancing the generalization performance of the LS-SVMR. However, the use of SSE as the loss function and the weight that is a function of the Euclidean distance between data points in the training set and query points resulted in a lack of robustness to outliers, especially those outliers close to query points. This is because the SSE is sensitive to outliers [25] as introduced previously, and in LWLS-SVMR, larger weights are given to points close to query points (i.e., small distances) and smaller weights are assigned to points far away from the query points (i.e., large distances). This formulation can lead LWLS-SVMR to produce a significantly biased prediction on query points which are nearby outliers (i.e., larger weights will be given to outliers, which significantly increases the contribution of outliers to prediction on query points).

Motivated by these existing solutions introduced previously, a novel, robust version of the local data-driven regression model, LWLS-SVMR, called RLWLS-SVMR,

is proposed in this paper to address the problem associated with non-robustness of LWLS-SVMR to input data corrupted by outliers. The proposed RLWLS-SVMR approach is validated according to its capability to broaden the application of LWLS-SVMR-based data-driven regression for cases where the input data may be contaminated by both non-extreme and extreme outliers. To be specific, three illustrative examples are given. First, simulated datasets with non-extreme and extreme outliers are provided for validating the performance of the proposed approach in univariate function approximation. Second, real-world multi-dimensional datasets are used to demonstrate the performance of the proposed method in multivariate function approximation for practical prediction in engineering disciplines. Third, the proposed approach is applied to extract the material properties for data-driven computational mechanics with a material dataset corrupted by outliers. For each example, the proposed approach is compared with existing relevant methods. The rest of this paper is organized as follows. Section 2 presents the methodology of the proposed RLWLS-SVMR, including the formulation of the RLWLS-SVMR and an iterative algorithm for RLWLS-SVMR. Section 3 presents the implementation procedure based on a hybrid algorithm of LWLS-SVMR and RLWLS-SVMR. Illustrative examples are given in Sect. 4. Finally, in Sect. 5, the conclusions are drawn.

2 Methodology

In this section, we present a novel robust data-driven regression approach, called robust LWLS-SVMR (RLWLS-SVMR), which is designed to overcome the fact that LWLS-SVMR is sensitive to outliers close to query points. The main difference between RLWLS-SVMR and LWLS-SVMR is the integration of an extra weight into the formulation of RLWLS-SVMR, which is a function of residuals and allows for iteration within the RLWLS-SVMR procedure, significantly reducing the negative interference of outliers. The major advantage of the proposed method over LWLS-SVMR is that it not only establishes the data-driven regression that is robust to input data contaminated by various types of outliers (i.e., non-extreme and extreme outliers) but also maintains the local nature, where, to predict a query point, the entire set of training data does not need to be fit. Instead, it only requires the fitting of a subset of training data nearby (relevant to) the query point. These characteristics yield a model that both overcomes the negative interference of outliers and avoids the potential negative influence of irrelevant points, achieving a suitable trade-off between the capacity of the learning system and the number of training data points. The detailed information regarding the mathematical equations and derivations for the formulation of the proposed RLWLS-SVMR is as follows.

2.1 Robust locally weighted least-squares support vector machines for regression (RLWLS-SVMR)

This section presents the mathematical formulation of the novel RLWLS-SVMR. Assume a multi-dimensional training set $\left\{\left(\boldsymbol{x}_i, y_i\right)\right\}_{i=1}^{n}$ is collected from a domain of interest and some observations (i.e., data points) have been corrupted by outliers. For the remainder of this paper, the following notations are utilized. Let R be the real numbers set; $\boldsymbol{x}_i \in R^p$ is a column vector with p dimensions (i.e., p variables) which can be written as $\boldsymbol{x}_i = \left(x_{i1}; \ldots ; x_{ip}\right)$, and $\boldsymbol{x}_i^T \in R^p$ represents the transpose of $\boldsymbol{x}_i$ and is a row vector with p dimensions which can be written as $\boldsymbol{x}_i^T = \left(x_{i1}, \ldots, x_{ip}\right)$, $y_i \in R$ is a real number; $\boldsymbol{X} \in R^{n\times p}$ is an $n \times p$ matrix which can be written as $\boldsymbol{X} = \left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)^T$; the training set $\left\{\left(\boldsymbol{x}_i, y_i\right)\right\}_{i=1}^{n}$ is an $n \times (p+1)$ matrix which includes n data points and each data point contains p explanatory variables (i.e., $\boldsymbol{x}_i \in R^p$) and one response (i.e., $y_i \in R$).

The mathematical formulation of the proposed RLWLS-SVMR is as follows:

(1) Given an independent test set $\left\{\left(\boldsymbol{x}_q, y_q\right)\right\}_{q=1}^{m}$ that is not included in the training set, for each query point $\boldsymbol{x}_q, q = 1, \ldots, m$, where the response value y_q is to be predicted and thus not considered in the following process.

(2) Define a subset $\left\{\left(\boldsymbol{x}_{(s)}, y_{(s)}\right)\right\}_{s=1}^{r}$ from the training set $\left\{\left(\boldsymbol{x}_i, y_i\right)\right\}_{i=1}^{n}$ by a parameter f_q, where f_q can take any value in the range $(0, 1]$; the number of data points in the subset is equivalent to $r = Ceil\left(f_q \times n\right)$, and the points in the subset are determined by the Euclidean distance metric via the following procedure:

   (a)) Calculate the Euclidean distance from each data point in the training set to each query point $\|\boldsymbol{x}_i - \boldsymbol{x}_q\|, i = 1, \ldots, n; q = 1, \ldots, m$, so for each query point, there is a distance vector $\boldsymbol{d}_q = \left(d_{q1}, \ldots, d_{qn}\right), q = 1, \ldots m$;

   (b)) Sort the entries in each distance vector increasingly, so a new sorted distance vector $\boldsymbol{d}_{(q)} = \left(d_{(q1)}, \ldots, d_{(qn)}\right), q = 1, \ldots m$ is obtained;

   (c)) The data points in the training set $\left\{\left(\boldsymbol{x}_i, y_i\right)\right\}_{i=1}^{n}$, corresponding to the first r entries in the sorted distance vector $\boldsymbol{d}_{(q)}$ (i.e., $d_{(q1)}, \ldots, d_{(qr)}$), can be selected as the subset $\left\{\left(\boldsymbol{x}_{(s)}, y_{(s)}\right)\right\}_{s=1}^{r}$. Note that, for different query points, the subset may vary.

(3) After the subset is determined, the learning objective of the proposed RLWLS-SVMR is to solve an optimization problem formulated by finding model parameters

$\boldsymbol{w} = \left(w_1; w_2; \ldots; w_h\right) \in R^h$ and $b \in R$, which is written as

$$\text{Minimize}: \; J\left(\boldsymbol{w}, e_s\right) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{1}{2}\gamma_q \sum_{s=1}^{r} v_q\left(\boldsymbol{x}_{(s)}\right)\beta_q\left(\boldsymbol{x}_{(s)}\right)e_s^2, q = 1, \cdots, m \tag{1}$$

$$\text{Subject to}: \; y_{(s)} = \boldsymbol{w}^T\varphi\left(\boldsymbol{x}_{(s)}\right) + b + e_s, \; s = 1, \cdots, r, \tag{2}$$

where $\gamma_q \in R, q = 1, \ldots, m$ is a regularization parameter; $e_s \in R, s = 1, \ldots, r$ is the error term; $\beta_q\left(\boldsymbol{x}_{(s)}\right), v_q\left(\boldsymbol{x}_{(s)}\right) \in R, s = 1, \ldots, r; q = 1, \ldots, m$ are weights that can take any value in the range$[\varepsilon, 1]$, $\beta_q\left(\boldsymbol{x}_{(s)}\right)$ is a function of the Euclidean distance where data points in a subset close to a query point have larger weights and far away from the query point have smaller weights; $v_q\left(\boldsymbol{x}_{(s)}\right)$ is a function of the residual where data points in a subset around the query point having large residuals have smaller weights and having small residuals have larger weights ;$\varepsilon \in R$ is a real number approaching 0; $\varphi\left(\boldsymbol{x}_{(s)}\right)$ is a feature vector, and $\varphi(\bullet): R^p \to R^h$ is a mapping function from p dimensions to a higher h-dimensional feature space.

Note: $\boldsymbol{x}_{(s)}$ is a column vector; thus, $\varphi\left(\boldsymbol{x}_{(s)}\right)$ is also a column vector.

If $\beta_q\left(\boldsymbol{x}_{(s)}\right)$ takes a value approaching ε, it means the point $\left(\boldsymbol{x}_{(s)}, y_{(s)}\right)$ is far away from the query point $\left(\boldsymbol{x}_q, y_q\right)$ (relatively large Euclidean distance) and plays a lesser role in the determination of y_q; while, if $\beta_q\left(\boldsymbol{x}_{(s)}\right)$ takes a value approaching one, it means the point $\left(\boldsymbol{x}_{(s)}, y_{(s)}\right)$ is close to the query point $\left(\boldsymbol{x}_q, y_q\right)$ (relatively small Euclidean distance) and plays an important role in the determination of y_q.

The Lagrangian function is established by the Lagrange multipliers method to solve Eqs. (1) and (2)

$$L\left(\boldsymbol{w}, b, e_s, \alpha_s\right) = J\left(\boldsymbol{w}, e_s\right) - \sum_{s=1}^{r}\alpha_s\left(\boldsymbol{w}^T\varphi\left(\boldsymbol{x}_{(s)}\right) + b + e_s - y_{(s)}\right), \tag{3}$$

where $\alpha_s \in R, s = 1, \ldots, r$ is a Lagrange multiplier (also called support values).

The Karush–Kuhn–Tucker (KKT) conditions for optimality are used by differentiating the variables in Eq. (3) above, which results in the following:

$$\begin{cases} \dfrac{\partial L}{\partial \boldsymbol{w}} = 0 \to \boldsymbol{w} = \sum_{s=1}^{r}\alpha_s\varphi\left(\boldsymbol{x}_{(s)}\right) \\ \dfrac{\partial L}{\partial b} = 0 \to 0 = \sum_{s=1}^{r}\alpha_s \\ \dfrac{\partial L}{\partial e_s} = 0 \to e_s = \dfrac{\alpha_s}{\gamma_q v_q\left(\boldsymbol{x}_{(s)}\right)\beta_q\left(\boldsymbol{x}_{(s)}\right)}, s = 1, \cdots, r; q = 1, \cdots, m \\ \dfrac{\partial L}{\partial \alpha_s} = 0 \to y_{(s)} = \boldsymbol{w}^T\varphi\left(\boldsymbol{x}_{(s)}\right) + b + e_s, s = 1, \cdots, r \end{cases}. \tag{4}$$

Rearranging Eq. (4) and eliminating $\boldsymbol{w}$ and e_s, using kernel function replace the inner product of feature vectors, the following matrix equation can be obtained:

$$\begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & K\left(\boldsymbol{x}_{(1)}, \boldsymbol{x}_{(1)}\right) + \frac{1}{\gamma_q v_q\left(\boldsymbol{x}_{(1)}\right)\beta_q\left(\boldsymbol{x}_{(1)}\right)} & K\left(\boldsymbol{x}_{(1)}, \boldsymbol{x}_{(2)}\right) & \cdots & K\left(\boldsymbol{x}_{(1)}, \boldsymbol{x}_{(r)}\right) \\ 1 & K\left(\boldsymbol{x}_{(2)}, \boldsymbol{x}_{(1)}\right) & K\left(\boldsymbol{x}_{(2)}, \boldsymbol{x}_{(2)}\right) + \frac{1}{\gamma_q v_q\left(\boldsymbol{x}_{(2)}\right)\beta_q\left(\boldsymbol{x}_{(2)}\right)} & \cdots & K\left(\boldsymbol{x}_{(2)}, \boldsymbol{x}_{(r)}\right) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & K\left(\boldsymbol{x}_{(r)}, \boldsymbol{x}_{(1)}\right) & K\left(\boldsymbol{x}_{(r)}, \boldsymbol{x}_{(2)}\right) & \cdots & K\left(\boldsymbol{x}_{(r)}, \boldsymbol{x}_{(r)}\right) + \frac{1}{\gamma_q v_q\left(\boldsymbol{x}_{(r)}\right)\beta_q\left(\boldsymbol{x}_{(r)}\right)} \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_r \end{bmatrix} = \begin{bmatrix} 0 \\ y_{(1)} \\ y_{(2)} \\ \vdots \\ y_{(r)} \end{bmatrix} \tag{5}$$

where $q = 1, \ldots, m$ and the kernel function is $K(\boldsymbol{x}_{(s)}, \boldsymbol{x}_{(t)}) = \varphi^T(\boldsymbol{x}_{(s)})\varphi(\boldsymbol{x}_{(t)}), s = 1, \ldots, r; t = 1, \ldots, r$.

For the determination of $\beta_q(\boldsymbol{x}_{(s)}) \in R, s = 1, \ldots, r; q = 1, \ldots, m$, for each query point $\boldsymbol{x}_q$, let $d_{(qr)}$ be the distance from $\boldsymbol{x}_q$ to the r^{th} nearest neighbor $\boldsymbol{x}_{(r)}$ [i.e., $d_{(qr)}$ is the maximum distance compared to $d_{(q1)}, \ldots, d_{(q(r-1))}$], and let $\beta_q(\boldsymbol{x}_{(s)}) = T\left(d_{(qr)}^{-1}\|\boldsymbol{x}_{(s)} - \boldsymbol{x}_q\|\right)$, where $T(\bullet)$ is a tri-cube weight function [56], which is defined as follows:

$$T(g) = f(x) = \begin{cases} \left(1 - |g|^3\right)^3, |g| < 1 \\ \varepsilon, |g| \geq 1 \end{cases}, \tag{6}$$

where ε can take any values close to 0, and in this work, $\varepsilon = 1e-4$ to avoid a zero in the denominator in Eq. (5).

The weight $v_q(\boldsymbol{x}_{(s)})$ in Eq. (5) is associated with the robustness to outliers close to a query point, and the determination of $v_q(\boldsymbol{x}_{(s)}) \in R, s = 1, \ldots, r; q = 1, \ldots, m$, for each query point $\boldsymbol{x}_q$ is discussed in detail in the next sub-section. The initial values of $v_q(\boldsymbol{x}_{(s)})$ for all points in the subset $\left\{(\boldsymbol{x}_{(s)}, y_{(s)})\right\}_{s=1}^{r}$ are set to one. Since the coefficient matrix in Eq. (5) is symmetric but not positive definite, it is difficult to directly solve Eq. (5) as there may be no inverse of the coefficient matrix which exists in the case where the coefficient matrix is close to singular. The method proposed by Suykens et al. [60] is used to solve this problem. After solving Eq. (5) [60], the Lagrange multiplier $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)$ and parameter b can be obtained, which can then be utilized to predict the query point $\boldsymbol{x}_q$ using the following:

$$\hat{y}(\boldsymbol{x}_q) = \sum_{s=1}^{r} \alpha_s K(\boldsymbol{x}_q, \boldsymbol{x}_{(s)}) + b. \tag{7}$$

The RBF kernel is utilized, which is defined as follows:

$$K(\boldsymbol{x}_q, \boldsymbol{x}_{(s)}) = exp\left(-\frac{\|\boldsymbol{x}_q - \boldsymbol{x}_{(s)}\|_2^2}{2\sigma_q^2}\right). \tag{8}$$

It should be noted that the proposed approach has three hyper-parameters that need to be tuned, while standard LS-SVMR with RBF kernels only require two tuning parameters. However, the hyper-parameter tuning process for the proposed approach is still straightforward. First, the parameter space for the additional tuning parameter (i.e., subset parameter $f_q \in (0, 1]$) is known, and thus, it is simple to incorporate the subset parameter f_q with the regularization and kernel parameters into the parameter optimization process using the grid search algorithm. Moreover, although the grid search algorithm requires more computational effort to optimize three parameters (rather than two), this problem can be effectively solved using a parallel scheme. This is because different combinations of these three hyper-parameters are evaluated independently, which allows these different combinations to be evaluated simultaneously rather than separately. This parallel scheme significantly enhances the computational efficiency.

2.2 Detection of negative effects due to outliers by proposed RLWLS-SVMR

As introduced in Sect. 2.1, the proposed RLWLS-SVMR is a robust *local* ML model. In this sense, all the points of the training set $\left\{(\boldsymbol{x}_i, y_i)\right\}_{i=1}^{n}$ are not necessarily considered in the training procedure for prediction of an individual query point $\boldsymbol{x}_q$. Considering the fact that the outliers are just a small portion of the entire training set, it is possible that outliers only exist in certain regions of the training set rather than distributed across the entire training set. In this case, the advantage of the proposed RLWLS-SVMR model is distinct. This is because, given a query point $\boldsymbol{x}_q$, the selected subset $\left\{(\boldsymbol{x}_{(s)}, y_{(s)})\right\}_{s=1}^{r}$ around the query point may not contain outliers, or the subset may contain outliers, but the outliers are sufficiently far away from the query point (see Fig. 1), such that the outliers have little negative effect on the prediction of the query point.

Figure 1 shows a schematic sketch illustrating how an outlier affects the prediction of a query point. Figure 1(a) shows the case where an outlier (red square point) exists in a selected subset $\left\{(\boldsymbol{x}_{(s)}, y_{(s)})\right\}_{s=1}^{r}$ but far away from the query point (black triangular point), while Fig. 1(b) shows the case where an outlier occurs close to the query point. Each point $(\boldsymbol{x}_{(s)}, y_{(s)})$ in this subset has a weight $\beta_q(\boldsymbol{x}_{(s)})$, and points close to the query point have larger $\beta_q(\boldsymbol{x}_{(s)})$ while points far away from the query point have smaller $\beta_q(\boldsymbol{x}_{(s)})$. In this way, points close to the query point have important contribution to the prediction of the query point, while those far away have little influence. If an outlier is far away from the query point, it is possible that the outlier will yield little negative influence on the prediction of the query point (see Fig. 1a) due to the smaller $\beta_q(\boldsymbol{x}_{(s)})$. This means that the weight $v_q(\boldsymbol{x}_{(s)}) = 1, s = 1, \ldots, r$ in Eq. (5) does not need to be updated, since the outlier does not cause significantly negative interference on the prediction. Thus, it is necessary to detect these types of negative effects, such that a computationally expensive iteration procedure to update the value of the weight $v_q(\boldsymbol{x}_{(s)})$ is not needed.

This can be achieved by selecting an appropriate region encompassing the query point (e.g., the region enclosed by the blue dashed rectangle in Fig. 1) by way of imposing a threshold. Then, the residuals between observed and predicted values within this region can be calculated, and a bound (positive number) can be selected. If the absolute average of the calculated residuals is smaller than the bound, it means that the outlier has little negative effect on prediction of the query point (e.g., Fig. 1a); however, if the absolute average is greater than the bound, the

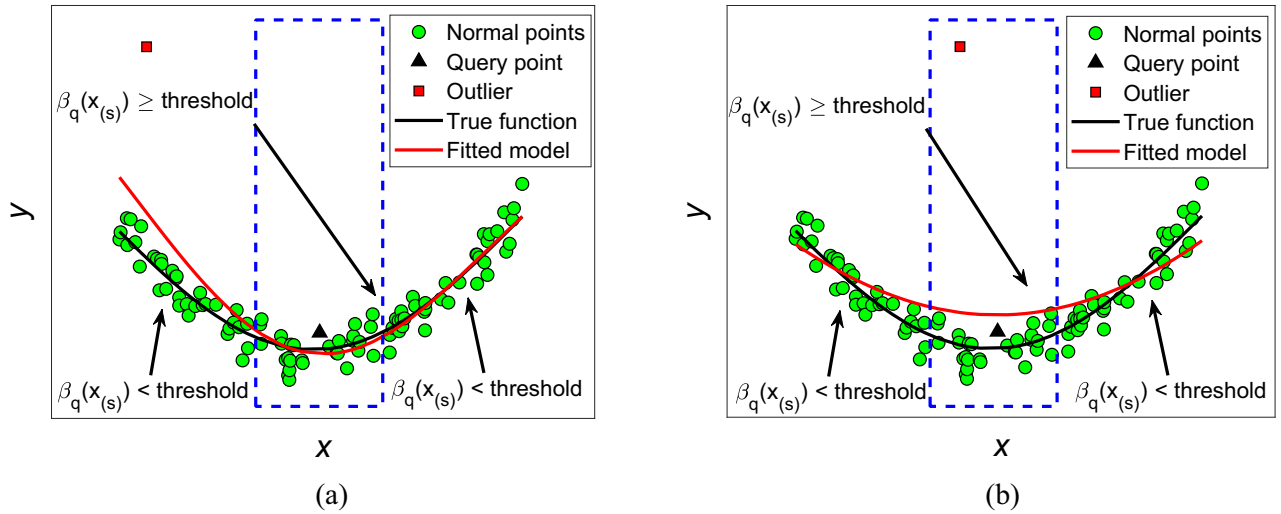

Fig. 1 Schematic sketch for detection of negative effects due to an outlier: **a** outlier far away from the query point has a diminished negative effect on prediction of the query point; **b** outlier close to the query point has a significantly negative effect on prediction of the query point. (Color figure online)

outlier is considered to have a sufficiently negative influence (e.g., Fig. 1b). The reason for choosing the absolute average of the residuals within the selected region as the judgment criterion is explained as follows. Considering the observation form $y_i = y_{itrue} + e_i$, if the proposed RLWLS-SVMR with weight $v_q(\boldsymbol{x}_{(s)}) = 1$ perfectly fits the true function, the predicted value will equal the true value ($\widehat{y}_i = y_{itrue}$). Thus, the residuals can be obtained by $y_i - \widehat{y}_i = y_i - y_{itrue} = e_i$. As e_i in classical statistical learning approaches is assumed zero mean [25], the absolute average of residuals within the selected region (i.e., the range within the blue dashed rectangle) will be zero, that is, $\left|E\left(\{e_i\}_{i=1}^{l}\right)\right| = 0$ (assume there are l data points within the blue dashed rectangle). The following algorithm based on the proposed RLWLS-SVMR is developed to realize this detection procedure:

Algorithm 1:

For each query point $\boldsymbol{x}_q, q = 1, \ldots, m$, **do**

(a) Given an optimal combination $(f_q, \gamma_q, \sigma_q^2)$, define a subset $\{(\boldsymbol{x}_{(s)}, y_{(s)})\}_{s=1}^{r}$ and weights $\beta_q(\boldsymbol{x}_{(s)})$ using Eq. (6);

(b) Set all weights $v_q(\boldsymbol{x}_{(s)})$ in Eq. (5) for the subset $\{(\boldsymbol{x}_{(s)}, y_{(s)})\}_{s=1}^{r}$ to 1;

(c) Solve Eq. (5) to obtain $\boldsymbol{\alpha}$, b, and compute residuals $e_s = \alpha_s / \left(\gamma_q v_q(\boldsymbol{x}_{(s)}) \beta_q(\boldsymbol{x}_{(s)})\right)$, where $s = 1, \ldots, r$;

(d) Set a threshold value and select the residuals within the region where the points having weights $\beta_q(\boldsymbol{x}_{(s)})$ are greater than the threshold;

In the above algorithm, flag = 1 represents the case when a negative influence is detected, while flag = 0 represents the opposite.

2.3 Robust regression by iterative RLWLS-SVMR

When outliers exist close to a query point $\boldsymbol{x}_q$ in the subset $\{(\boldsymbol{x}_{(s)}, y_{(s)})\}_{s=1}^{r}$, the response value predicted by the proposed RLWLS-SVMR with $v_q(\boldsymbol{x}_{(s)}) = 1, s = 1, \ldots, r$, for the query point $\boldsymbol{x}_q$ will be negatively affected by those outliers (see Fig. 1b). Thus, the *algorithm 2* is developed here to eliminate the negative influence of outliers based on the proposed RLWLS-SVMR by iteratively updating the weight $v_q(\boldsymbol{x}_{(s)})$, as a function of e_s. These weights are computed via Eq. (9) and according to Suykens et al. [49]

$$v_q\left(\boldsymbol{x}_{(s)}\right)=\begin{cases}1 & if\left|e_s/\delta\right|\le c_1\\ \frac{c_2-\left|e_s/\delta\right|}{c_2-c_1} & if c_1\le\left|e_s/\delta\right|\le c_2\\ \varepsilon & otherwise\end{cases}, \tag{9}$$

where $c_1=2.5$, $c_2=3$, $\varepsilon=10^{-4}$, and $\delta=1.483MAD\left(\left\{e_s\right\}_{s=1}^{r}\right)$ is a robust estimate where MAD is the median absolute deviation and other variables are defined previously.

After the calculation of $v_q\left(\boldsymbol{x}_{(s)}\right)$ is carried out, the iterative RLWLS-SVMR to predict the response value of a query point $\boldsymbol{x}_q$ is achieved by the following procedure:

(e) Set a bound value and calculate the absolute of average of the selected residuals, and compare the absolute and bound;
If absolute > bound **then**
Flag = 1
else
Flag = 0
end if
end for

For each query point $\boldsymbol{x}_q, q=1,\ldots,m$, **do**
1. Initialization stage:
(a) Given an optimal combination $\left(f_q,\gamma_q,\sigma_q^2\right)$, define a subset $\left\{\left(\boldsymbol{x}_{(s)},y_{(s)}\right)\right\}_{s=1}^{r}$ and weights $\beta_q\left(\boldsymbol{x}_{(s)}\right)$ using Eq. (6);
(b) Set all weights $v_q\left(\boldsymbol{x}_{(s)}\right), s=1,\ldots,r$ in Eq. (5) for the subset $\left\{\left(\boldsymbol{x}_{(s)},y_{(s)}\right)\right\}_{s=1}^{r}$ to 1;
(c) Solve Eq. (5) to obtain $\boldsymbol{\alpha}$, b, and compute $e_s=\alpha_s/\left(\gamma_q v_q\left(\boldsymbol{x}_{(s)}\right)\beta_q\left(\boldsymbol{x}_{(s)}\right)\right)$, where $s=1,\ldots,r$.
2. Iterative stage:
Set the maximum iterative number N, tolerance tol, count $i=0$, and $t=Inf$
while $t>tol$ && $i<N$ **do**
(a) Set $\boldsymbol{\alpha}^{(i)}=\boldsymbol{\alpha}$, $b^{(i)}=b$, $e_s^{(i)}=e_s$, and $v_q^{(i)}\left(\boldsymbol{x}_{(s)}\right)=v_q\left(\boldsymbol{x}_{(s)}\right), s=1,\ldots,r$;
(b) Compute the robust estimate $\delta^{(i)}=1.483MAD\left(\left\{e_1^{(i)},\ldots,e_r^{(i)}\right\}\right)$;
(c) Update the weights $v_q^{(i+1)}\left(\boldsymbol{x}_{(s)}\right)$ from $\delta^{(i)}$ and $e_s^{(i)}$ using Eq. (9);
(d) Solve Eq. (5) to obtain the $\boldsymbol{\alpha}^{(i+1)}$ and $b^{(i+1)}$;
(e) Update the $e_s^{(i+1)}=\alpha_s^{(i+1)}/\left(\gamma_q v_q^{(i+1)}\left(\boldsymbol{x}_{(s)}\right)\beta_q\left(\boldsymbol{x}_{(s)}\right)\right), s=1,\ldots r$;
(f) Calculate $t=\left\|\boldsymbol{\alpha}^{(i+1)}-\boldsymbol{\alpha}^{(i)}\right\|$;
(g) Set $\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(i+1)}, b=b^{(i+1)}, e_s=e_s^{(i+1)}$, and $v_q\left(\boldsymbol{x}_{(s)}\right)=v_q^{(i+1)}\left(\boldsymbol{x}_{(s)}\right), s=1\ldots,r$;
(h) Set $i=i+1$
end while
3. Output stage:
(a) Output the final $\boldsymbol{\alpha}$ and b from the procedure 2
(b) Given $\boldsymbol{\alpha}$ and b, predict the response value $\hat{y}_q$ of the query point $\boldsymbol{x}_q$ using Eq. (7).
end for

Since there is no weight function $v_q\left(\boldsymbol{x}_{(s)}\right)$ in the LWLS-SVMR, the obtained Lagrange multiplier $\boldsymbol{\alpha}$ and parameter b cannot be updated. If outliers surround the query points, the obtained $\boldsymbol{\alpha}$ and b will be negatively affected, resulting in the final predictive model that is negatively influenced by outliers. However, from Algorithm 2, it can be observed that, compared to LWLS-SVMR, the Lagrange multiplier $\boldsymbol{\alpha}$ and parameter b for the proposed RLWLS-SVMR can be updated in each iteration due to the update of the weight $v_q\left(\boldsymbol{x}_{(s)}\right)$ until convergence is reached. At this time, the finally updated $\boldsymbol{\alpha}$ and b can result in a predictive model (i.e., Eq. 7) that is robust to both non-extreme and extreme outliers. This is because, in each iteration, the negative effect induced by non-extreme and extreme

outliers is reduced by the updated weight $v_q(\boldsymbol{x}_{(s)})$, which in turn updates the predictive model by updating the Lagrange multiplier $\boldsymbol{\alpha}$ and parameter b. When convergence is reached, the negative effect from non-extreme and extreme outliers is almost eliminated, and thus, the finally updated predictive model is not affected by the outliers.

3 Implementation of a hybrid algorithm for proposed RLWLS-SVMR

This section introduces the implementation procedure of the proposed RLWLS-SVMR using a hybrid algorithm. As introduced in Sect. 2.2, outliers are only representative of a small amount of the training data, and therefore, not all of the regions will necessarily contain outliers. It is true that some query points may be far away from outliers. In this case, the negative effect from outliers for the predictions of these query points can be ignored, and the results predicted by LWLS-SVMR [i.e., the proposed RLWLS-SVMR with the weight $v_q(\boldsymbol{x}_{(s)}) = 1, s = 1, \ldots, r$] can be trusted. By combining detection of the negative effect of outliers and the iterative version of RLWLS-SVMR, an efficient hybrid algorithm is developed to predict query points by adaptively using either LWLS-SVMR or iterative version of RLWLS-SVMR depending on whether or not a negative effect is detected. The hybrid algorithm is implemented in this paper as the following:

Hybrid algorithm:

For each query point $\boldsymbol{x}_q, q = 1, \ldots, m$, **do**
  Given an optimal combination $(f_q, \gamma_q, \sigma_q^2)$, detect if there is any negative influence induced by outliers using *Algorithm 1*
  If flag = 0 **then**
    Predict the response $\hat{y}_q$ of the query point $\boldsymbol{x}_q$ according to $\boldsymbol{\alpha}$ and b obtained in *Algorithm 1* using Eq. (7) and record the predicted result;
  else
    Perform an iterative procedure using *Algorithm 2* and record the final predicted result;
  end if
end for

In addition to the implementation of RLWLS-SVMR, other relevant ML approaches are also implemented for performance comparison. The relevant ML approaches are LS-SVMR [35], weighted LS-SVMR (WLS-SVMR) [49], and iterative WLS-SVMR (IWLS-SVMR) [50]. Note that the LWLS-SVMR is already incorporated into the hybrid algorithm and the disadvantage of LWLS-SVMR for training input datasets corrupted by outliers is already discussed in theory (Sect. 2.2). Thus, direct implementation of LWLS-SVMR is not included in this study. The LS-SVMR serves as the baseline, since other models used in this study are all variants of LS-SVMR, to address the problems associated with input datasets corrupted by outliers. The main difference between the proposed RLWLS-SVMR and WLS-SVMR and IWLS-SVMR is that the RLWLS-SVMR is a robust, *local* model, whereas both WLS-SVMR and IWLS-SVMR are robust, *global* models. The difference between robust local and global models has been introduced in Sects. 1 and 2. The detailed formulations for LS-SVMR, WLS-SVMR, and IWLS-SVMR can be found in the original references [35, 49, 50]. The RBF kernel is also applied for both LS-SVMR, WLS-SVMR, and IWLS-SVMR. The optimal hyper-parameter combinations for all four models are obtained using fivefold cross-validation on the training data [61].

4 Numerical experiments

This section presents three illustrative examples for validating the proposed approach. First, to assess the proposed approach for a dataset $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ corrupted by outliers, we present two examples using both simulated and multi-dimensional real-world datasets. The proposed method is compared with LS-SVMR, WLS-SVMR, and IWLS-SVMR for all these two examples. Then, the proposed approach is applied for data-driven computational elasticity with a material dataset corrupted by outliers. The generalization performances for simulated datasets are quantified by the coefficient of determination (R^2) (Eq. 10), mean absolute error (MAE) (Eq. 11), and root-mean-square error (RMSE) (Eq. 12). Since R^2, MAE, and RMSE are sensitive to outliers in the test set, and the test set in the real-world datasets may contain outliers (i.e., we never know the true value in the real-world dataset, but we do know the true value in the simulated dataset), the performance

for real-world datasets is quantified by a robust variant of R^2 (R_R^2) (Eq. 13) [62], which has been successfully demonstrated as robust to outliers in the test set [63, 64]. Given response variable $\boldsymbol{y} = \{y_i\}_{i=1}^n$ and predicted response $\hat{\boldsymbol{y}} = \{\hat{y}_i\}_{i=1}^n$, R^2, MAE, RMSE, and R_R^2 are calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^n \left(y_i - \bar{y}\right)^2} \tag{10}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left|y_i - \hat{y}_i\right| \tag{11}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n \left(y_i - \hat{y}_i\right)^2}{n}} \tag{12}$$

$$R_R^2 = 1 - \left(\frac{median\left(\left|\boldsymbol{y} - \hat{\boldsymbol{y}}\right|\right)}{mad(\boldsymbol{y})}\right)^2, \tag{13}$$

where $mad(\boldsymbol{y}) = median(|\boldsymbol{y} - median(\boldsymbol{y})|)$ is the median absolute deviation of $\boldsymbol{y}$.

Both the original and robust variants of R^2 are typically in the range of [0, 1] with 1 representing a perfect prediction. However, in some cases, the R^2 could be negative and a negative R^2 value corresponds to extremely poor prediction, which means that the model breaks down. Both MAE and RMSE values will be equal to or greater than 0, with 0 representing perfect prediction. We now use a very simple example to illustrate that original R^2, MAE, and RMSE are sensitive to outliers in the test set, but the robust variant of R^2 is robust to outliers in the test set. Assume a response variable in the test set is corrupted by one outlier, $\boldsymbol{y} = (2, 4, 6, 8, 100, 12, 14)$ [i.e., the fifth element (100) is corrupted, and the actual value is 10]. A robust model is applied to predict the response values for the test set, and the predicted response is $\hat{\boldsymbol{y}} = (2, 4, 6, 8, 10, 12, 14)$, which means that the robust model perfectly predicts the response in the test set. However, if we use the statistical indicators above (i.e., Eq. 10–13) to quantify the performance of the robust model, one can obtain the performance of this robust model is − 0.09, 12.86, 34.01, and 1, respectively. Therefore, only the robust variant of R^2 reflects the actual performance of the robust model, and the other statistics are sensitive to outliers and fail to quantify the actual performance. Note that if more outliers exist in the test set, the robust variant of R^2 may also fail to reflect the actual performance, but it is still more robust than RMSE, MAE, and the original R^2 [62].

4.1 Example 1: simulated datasets

In this example, we generate four synthetic datasets corrupted by four combinations of two different types of random error terms and two different types of outliers to show the robustness of RLWLS-SVMR. In the real-world, the random error term reflects the data noise that cannot be avoided (note that noise is not necessarily representative of an outlier [25]) as purely clean data are impossible [25]. The error model proposed by [65] is used to generate these four synthetic datasets. Specifically, the random error terms (i.e., noise) are simulated using Gaussian distribution of zero mean and either constant or non-constant variance. The outliers are simulated by either a Gaussian distribution with higher variance or a standard Cauchy distribution with heavy tails. In this setting, a dataset $\left\{\left(\boldsymbol{x}_i, y_i\right)\right\}_{i=1}^n$ not corrupted by outliers is simulated from a *sinc* function, which is defined in this way

$$y_i = \frac{sin\left(x_i\right)}{x_i} + e_i,$$

where x_i is drawn from a uniform distribution $x_i \sim U[-10, 10]$, and e_i is a random error term that is drawn from a Gaussian distribution using either constant variance, i.e., $e_i \sim N\left(0, 0.01^2\right)$ or non-constant variance, i.e., $e_i \sim N\left(0, \sigma_i^2\right)$ and $\sigma_i \sim U[0.01, 0.05]$. We select the smaller variance to distinguish the noise from outliers in the regression setting (*Fig. 2*).

The number of normal data points following the definition above is 162. Another 38 points are defined as the potential outliers, where e_i is drawn from either a Gaussian distribution with higher variance, i.e., $e_i \sim N\left(0, 1^2\right)$ or a standard Cauchy distribution with heavy tails, i.e., $e_i \sim C(0, 1)$. A total of 200 data points, serving as the training data, are drawn from the mixture procedure introduced above. By setting different random number seeds, four combinations of error terms and outliers are performed to form four synthetic training datasets where the locations of outliers differ to more extensively evaluate the robustness of these four ML models, as shown in Fig. 2 (a, c, e, g). In Fig. 2, the four synthetic training datasets are shown in the left subfigures (a, c, e, g) which differ according to the error and outlier distributions as follows, while the corresponding test sets are shown in the right subfigures (b, d, f, h): (a, b) *Synthetic 1:* the error terms for normal points are drawn by $e_i \sim N\left(0, 0.01^2\right)$ and the potential outliers are drawn by $e_i \sim N\left(0, 1^2\right)$; (c, d) *Synthetic 2:* the error terms for normal points are drawn by $e_i \sim N\left(0, \sigma_i^2\right)$ and $\sigma_i \sim U[0.01, 0.05]$, and the potential outliers are drawn by $e_i \sim N\left(0, 1^2\right)$; (e, f) *Synthetic 3*: the error terms for normal

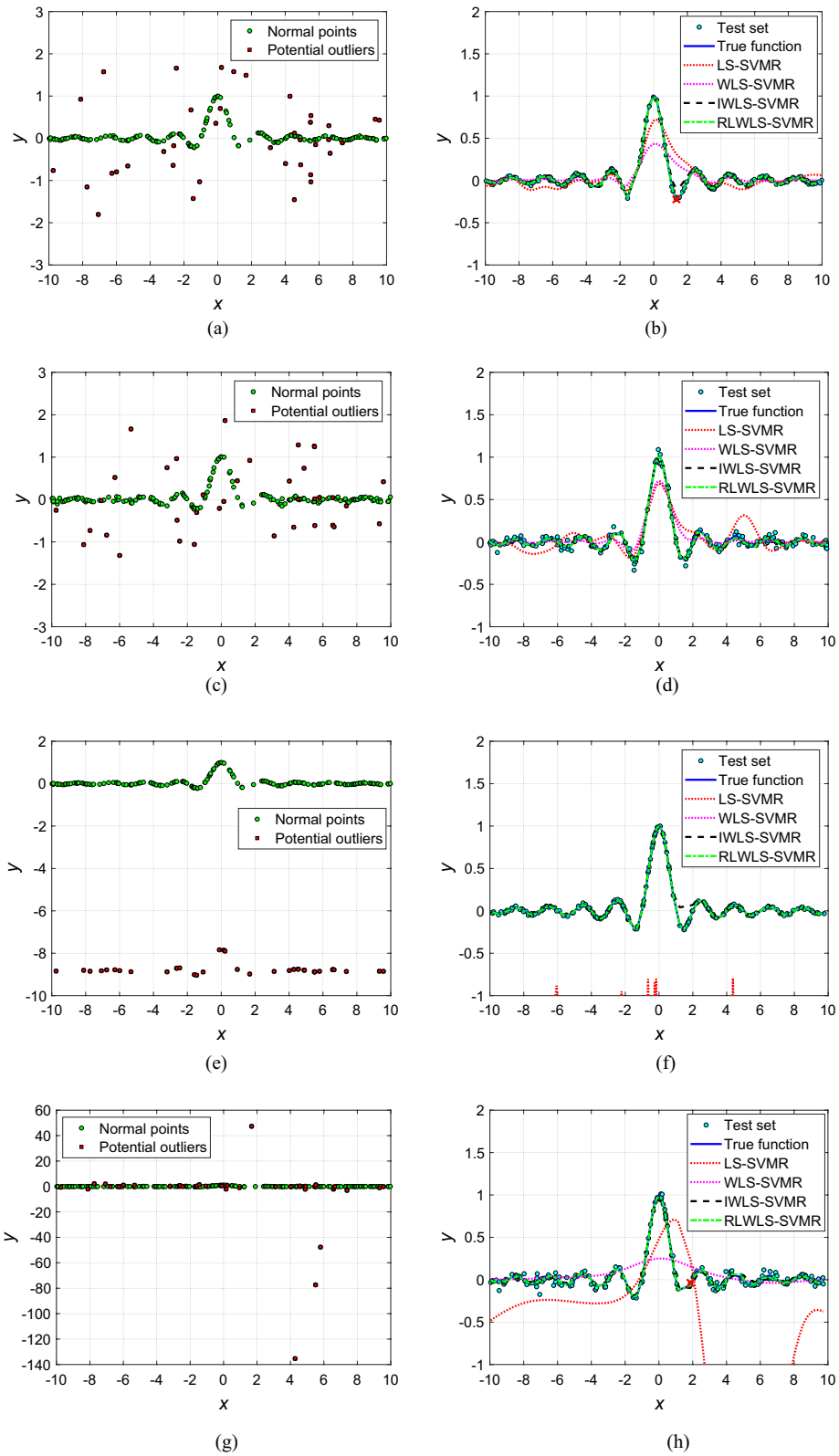

Fig. 2 Left subfigures (**a**, **c**, **e**, **g**): training of a *sinc* function with four synthetic training datasets (with various error and simulated outlier characteristics employed to plague the training data). Right subfigures (**b**, **d**, **f**, **h**): testing (estimation of the *sinc* function) by LS-SVMR, WLS-SVMR, IWLS-SVMR, and RLWLS-SVMR. (Color figure online)

points are drawn by $e_i \sim N\left(0, 0.01^2\right)$ and the potential outliers are drawn by $e_i \sim C(0, 1)$; and (g, h) *Synthetic 4:* the error terms for normal points are drawn by $e_i \sim N\left(0, \sigma_i^2\right)$ and $\sigma_i \sim U[0.01, 0.05]$ and the potential outliers are drawn

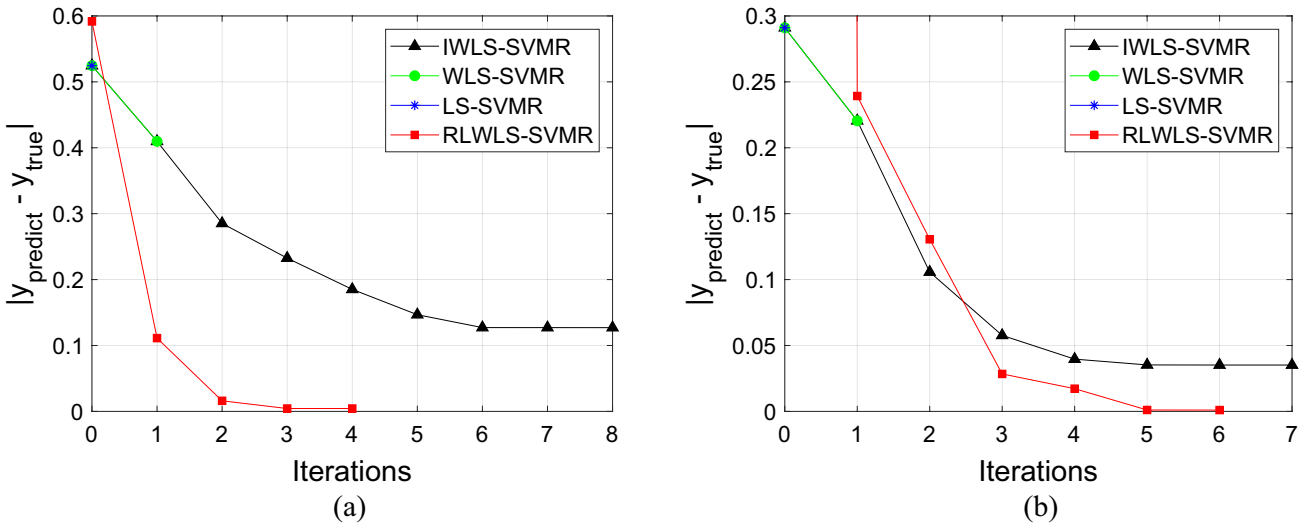

Fig. 3 The comparison of results in terms of predicted absolute error versus number of iterations for the selected two points (i.e., red 'X' points in Fig. 2b and h) between LS-SVMR, WLS-SVMR, IWLS-SVMR, and proposed RLWLS-SVMR. **a** Non-extreme outlier case: results for the data point denoted by a red 'X' in Fig. 2b. **b** Extreme outlier case: results for the data point denoted by a red 'X' in Fig. 2h. (Color figure online)

by $e_i \sim C(0,1)$. Note that the potential outliers are only applied to the four synthetic training datasets. It is clearly observed that not all of the potential outliers are real outliers, and only the points far from the bulk of the data points are true outliers (i.e., y-outliers). Another 200 independent test data points [i.e., Fig. 2 (b, d, f, h)] not corrupted by outliers (i.e., there are no outliers in the test data) corresponding to four different synthetic training datasets are drawn to test the performance of data-driven regression constructed by LS-SVMR, WLS-SVMR, IWLS-SVMR, and the proposed RLWLS-SVMR. The scatter plots of training and test data as well as the predictions on the test data by LS-SVMR, WLS-SVMR, IWLS-SVMR, and RLWLS-SVMR are presented in Fig. 2.

It should be noted that for LS-SVMR, WLS-SVMR, and IWLS-SVMR, a global model is formed using the entire training dataset before predicting the query points in the test dataset. For the proposed RLWLS-SVMR, different query points in the test dataset (i.e., points in the test dataset are also query points to be predicted) are predicted by distinct, individual local models formed by training different subsets of training data to achieve trade-off between prediction capacities of learning systems and number of training data for different query points. A comparison of the results between LS-SVMR, WLS-SVMR, IWLS-SVMR, and the proposed RLWLS-SVMR on the four test datasets is shown in Fig. 2(b, d, f, h). By observation, compared to the true function, LS-SVMR is negatively affected by outliers, especially by those produced by the standard Cauchy distribution with heavy tails (i.e., extreme outliers), where the LS-SVMR is influenced heavily in the direction of outliers, leading to the significant deviation from the true function (Figs. 2f and h). The WLS-SVMR improves the performance of LS-SVMR but still suffers negative effects. By contrast, both IWLS-SVMR and the proposed RLWLS-SVMR perform much more robustly to both non-extreme and extreme outliers, where both methods overcome the negative interference from outliers and very closely fit the true function.

To show how the proposed approach works under the presence of non-extreme and extreme outliers in the training datasets, two points in the test datasets (i.e., red 'X' points in Figs. 2b and h) are selected to explicitly investigate the relation between the predicted absolute error (i.e., $\left|y_{predict} - y_{true}\right|$) and number of iterations. The reason to select these two points is because the location of one point (i.e., red 'X' points in the Figs. 2b) is next to a non-extreme outlier in the training dataset (see Fig. 2a), and the location of another point (i.e., red 'X' points in the Figs. 2h) is close to an extreme outlier in the training dataset (see Fig. 2g). This strategy can clearly show how the proposed method reduces the negative effect from non-extreme and extreme outliers and further emphasizes how the proposed approach differs from the existing relevant approaches. The comparisons of results between LS-SVMR, WLS-SVMR, IWLS-SVMR, and proposed RLWLS-SVMR are reported in Fig. 3. By observation of Fig. 3, when the number of iterations is equal to zero, the proposed RLWLS-SVMR produces the largest error under the presence of non-extreme (see Fig. 3a) and extreme (see Fig. 3b) outliers. This is because at this time, the weight

Table 1 Performance comparison between LS-SVMR, WLS-SVMR, IWLS-SVMR, and RLWLS-SVMR in terms of original R^2, RMSE, and MAE

Datasets	Models	RMSE	MAE	R^2
Synthetic dataset 1	LS-SVMR	0.1163	0.0742	0.5182
	WLS-SVMR	0.1115	0.0589	0.5576
	IWLS-SVMR	0.0213	0.0070	0.9839
	RLWLS-SVMR	**0.0052**	**0.0040**	**0.9990**
Synthetic dataset 2	LS-SVMR	0.1380	0.0992	0.5834
	WLS-SVMR	0.0847	0.0515	0.8428
	IWLS-SVMR	0.0137	0.0106	0.9959
	RLWLS-SVMR	**0.0083**	**0.0051**	**0.9985**
Synthetic dataset 3	LS-SVMR	1.7270	1.6301	-43.5127
	WLS-SVMR	1.7160	1.6964	-42.9457
	IWLS-SVMR	0.0490	0.0164	0.9642
	RLWLS-SVMR	**0.0019**	**0.0011**	**0.9999**
Synthetic dataset 4	LS-SVMR	1.6499	1.0328	-42.4528
	WLS-SVMR	0.1997	0.1062	0.3633
	IWLS-SVMR	0.0229	0.0179	0.9916
	RLWLS-SVMR	**0.0085**	**0.0050**	**0.9989**

The synthetic datasets represent the training data corrupted by outliers and the original R^2, RMSE, and MAE are computed on corresponding test datasets between predicted and true values

The bold values represent the best performance

$v_q\left(\boldsymbol{x}_{(s)}\right) = 1$ and is not updated (see steps 1b and c in Algorithm 2 for more details), which results in the reversion of the proposed RLWLS-SVMR to LWLS-SVMR as introduced in Sect. 2. Since the outliers are close to the selected two points, the LWLS-SVMR gives larger weights to outliers, which enhances the negative effect of outliers and causes the largest error. However, when the number of iterations increases [i.e., the weight $v_q\left(\boldsymbol{x}_{(s)}\right)$ is updated], the error produced by the proposed method decreases until it is no longer reduced (i.e., convergence is reached). This is reflected by the red 'line + rectangular points' in Fig. 3a and b. At this time, the proposed RLWLS-SVMR almost eliminates the negative effect induced by outliers and produces the predicted values that are close to the true values for the selected two points. Additionally, in comparison with the three global models (i.e., LS-SVMR, WLS-SVMR, and IWLS-SVMR), the proposed RLWLS-SVMR yields the best performance, as shown in Fig. 3. This is because these three global models require the entire training dataset to be used for predicting the selected two points, while the proposed RLWLS-SVMR only requires the subsets nearby (or relevant to) the selected two points as introduced in Sect. 2. Therefore, the proposed approach further improves the performance of global models by both overcoming the negative interference of outliers and avoiding the potential negative influence of irrelevant points, achieving a suitable trade-off between the capacity of the learning system and the number of training data points.

Table 1 presents the metrics of original R^2, RMSE, and MAE for LS-SVMR, WLS-SVMR, IWLS-SVMR, and RLWLS-SVMR in terms of test datasets. Since these datasets are simulated and we know the true values of them, these metrics can give correct quantifications for the actual performance of these four ML models. Thus, it can be concluded that both IWLS-SVMR and RLWLS-SVMR do adequately capture the true function, and the proposed RLWLS-SVMR has the highest R^2 and lowest RMSE and MAE values, which deem it as the best model for these types of datasets among the four ML models.

4.2 Example 2: multi-dimensional real-world datasets

To further investigate the robustness of the proposed RLWLS-SVMR for multi-dimensional problems and demonstrate its practical application in engineering, we employ four multi-dimensional real-world engineering datasets to test the model performance and compare it with LS-SVMR, WLS-SVMR, and IWLS-SVMR. These eight benchmark datasets (and associated tasks) are the following: (1) *Reinforced concrete (RC) columns* (predicting the lateral strength) [66]; (2) *Automobile characteristics* (predicting the fuel consumption) [67]; (3) *Servo* (predicting the rising time of a servomechanism) [67]; and (4) *Nelson* (predicting the dielectric breakdown strength) [68]. The detailed information for all four real-world datasets can be found in the provided websites in the references. The final results are reported for all four datasets to demonstrate the broad application of the proposed approach in solving various engineering problems, and a detailed discussion of how the models perform is carried out for the RC column dataset to thoroughly explain the proposed approach and its performance.

Accurate modeling of lateral strength of RC columns is a very important topic in structural and earthquake engineering, as the lateral strength is an important factor for the design of buildings [76, 77]. In this specific example, we test the prediction performance of LS-SVMR, WLS-SVMR, IWLS-SVMR, and the proposed RLWLS-SVMR on lateral strength prediction of RC columns. A database including 160 RC circular columns is utilized. This database is extracted from the PEER Structural Performance Database compiled by Berry et al. [66]. The input predictors (i.e., explanatory variables) are column gross sectional area (X_1), concrete compressive strength (X_2), column cross-sectional effective depth (X_3), longitudinal reinforcement yield stress (X_4) and area (X_5), transverse reinforcement yield stress (X_6)

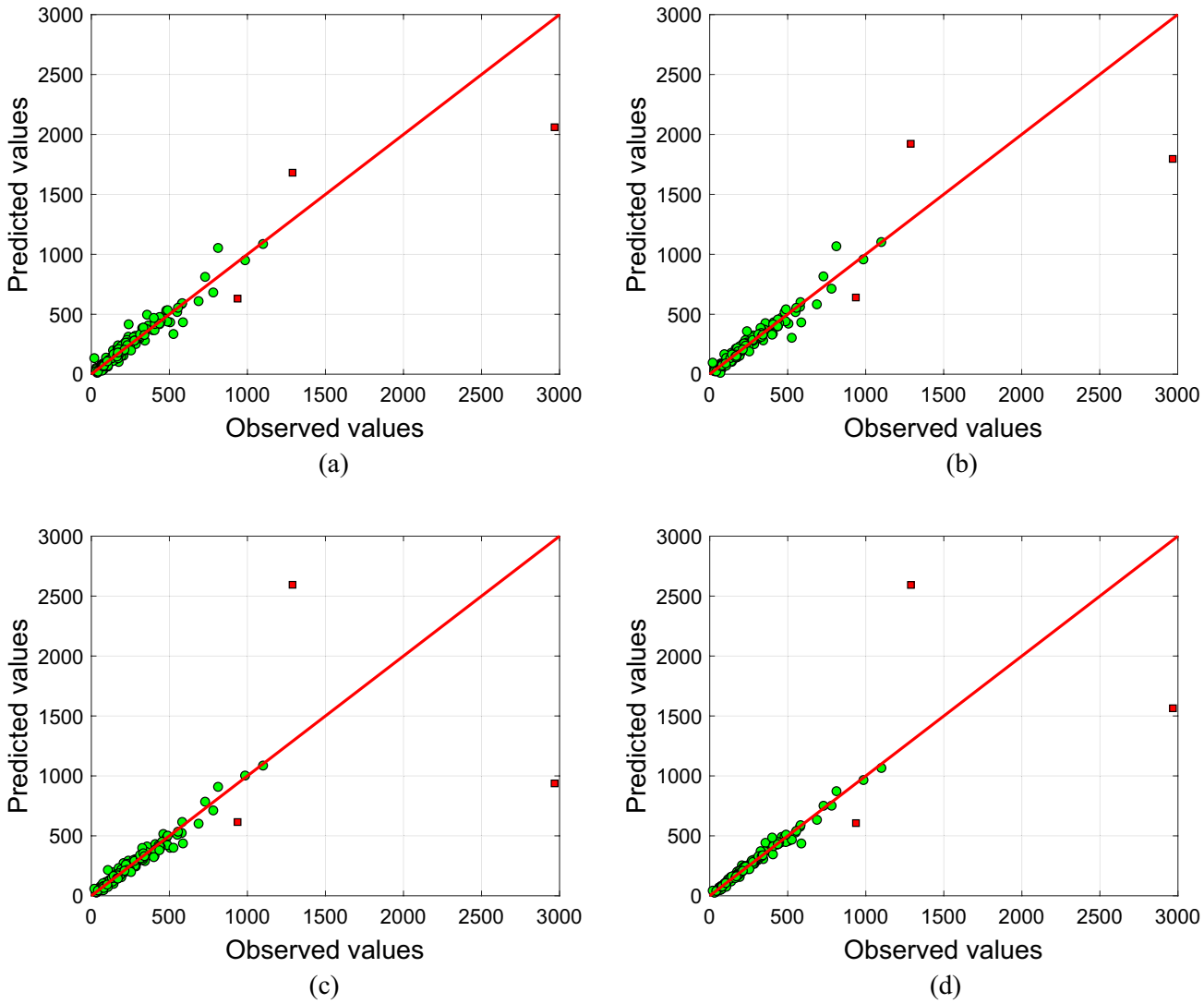

Fig. 4 Comparison of results using leave-one-out (LOO) cross-validation procedure on 160 RC columns of: **a** LS-SVMR, **b** WLS-SVMR, **c** IWLS-SVMR, and **d** RLWLS-SVMR. (Color figure online)

and area (X_7), stirrup spacing to effective depth ratio (X_8), shear span to effective depth ratio (X_9), and applied axial load (X_{10}), and the response variable is lateral strength (y), defined at the maximum shear force (kN) in the hysteretic force–deformation curve. Detailed information regarding this dataset can be found in Berry et al. [66].

We use a leave-one-out (LOO) cross-validation procedure [24] to test the performance of LS-SVMR, WLS-SVMR, IWLS-SVMR, and RLWLS-SVMR on lateral strength prediction of these 160 RC columns as well as for the other seven real-world datasets. The performance of these ML models on prediction in these eight real-world datasets is quantified by the robust variant of R^2 defined in Eq. (13). Note that the true values of the response variables in the real-world datasets are unknown. This is because the observed value of the response variables in real-world datasets contains a random error term (i.e., $y = y_{true} + e$), and the random error is unknown. If outliers exist in the real-world dataset, the original R^2, RMSE, and MAE will be sensitive to these outliers and fail to reflect the prediction performance of these four ML models based on the LOO cross-validation procedure, while the robust variant of R^2 is more robust to outliers and can give a more objective evaluation, as discussed previously. Additionally, it is worth noting that a robust estimator is able to detect outliers where points possess large residuals from the robust estimation, while a non-robust estimator cannot be used for this purpose, because the outliers may possess very small residuals [25].

A comparison of results is presented in Fig. 4. By observation of Fig. 4, the green points in all four ML models flock around the red lines which indicates that the predicted and observed values are equal (i.e., perfect prediction). However,

Table 2 Performance comparison between LS-SVMR, WLS-SVMR, IWLS-SVMR and RLWLS-SVMR on four benchmark real-world engineering datasets in terms of the robust variant of R^2 using LOO cross-validation procedure

Datasets	Number of observations	Number of predictors	LS-SVMR	WLS-SVMR	IWLS-SVMR	RLWLS-SVMR
Columns	160	10	0.9747	0.9756	0.9837	**0.9928**
Auto MPG	392	7	0.9393	0.9427	0.9434	**0.9723**
Servo	167	4	0.7367	0.8326	0.8789	**0.9265**
Nelson	128	2	0.8626	0.8657	0.8675	**0.9012**

The bold values represent the best performance

compared to IWLS-SVMR (Fig. 4c) and RLWLS-SVMR (Fig. 4d), the green points in LS-SVMR (Fig. 4a) and WLS-SVMR (Fig. 4b) are much more scattered. Additionally, there are three red square points in all four ML models which are distant from the red lines. Compared to LS-SVMR and WLS-SVMR, the two red points (i.e., values more than 1000 kN in the observed value direction in Fig. 4) in IWLS-SVMR and RLWLS-SVMR are much further from the red lines, which lead to higher residuals (i.e., difference between observed and predicted values). The other remaining red point (i.e., value less than 1000 kN in the observed value direction in Fig. 4) appears to maintain nearly the same deviation in all four ML models (i.e., the residuals for this red point in all four ML models are almost equivalent).

By analysis of the dataset, it is found that these two red points (i.e., values more than 1000 kN in the observed value direction in Fig. 4) correspond to two full-scale column tests conducted by Stone and Cheok [69], where the section dimensions (explanatory variable) and lateral strength (response variable) of these two columns are extreme values which are far larger than all other remaining columns in the dataset. It is also found that the other remaining red point corresponds to a column test performed by Priestley et al. [70] where the applied axial load (explanatory variable) on this column is an extreme value which is much larger than all other columns in the dataset. Thus, these three red points are detected and identified as high leverage points (i.e., extreme values in the x direction; note that this does not take y into account, and if a high leverage point is also an outlier, it will negatively affect the performance of a non-robust estimator [25]). By observation of Fig. 4, it is evident that the LS-SVMR is influenced heavily in the direction of these two high leverage points [i.e., values more than 1000 kN (outliers)]. This negative effect for LS-SVMR is exhibited by smaller residuals for the two high leverage points (outliers) but greater scatter in the remaining points than the results for WLS-SVMR, IWLS-SVMR, and RLWLS-SVMR. The WLS-SVMR slightly reduces the negative interference from these points where the residuals are slightly larger, and the green points are slightly less scattered in comparison to LS-SVMR. However, both IWLS-SVMR and RLWLS-SVMR improve the prediction on green points by significantly reducing the negative interference, where the green points are much less scattered and those two red points are far away from the red lines. The proposed RLWLS-SVMR performs better than IWLS-SVMR where the green points in RLWLS-SVMR are less scattered than those in IWLS-SVMR. Since the other remaining red point does not deleteriously change the prediction for all four ML models, it can be concluded that this leverage point is a good leverage point, while the other two red points mentioned above are bad leverage points that are also outliers. The final results for the RC column dataset as well as for the other seven datasets mentioned previously are reported in Table 2. From Table 2, it is observed that the proposed RLWLS-SVMR performs best across all eight benchmark real-world datasets.

4.3 Example 3: computational mechanics application

In computational mechanics, Kirchdoerfer and Ortiz [17] introduced the methodology of data-driven computational mechanics, where the traditional constitutive equations are substituted by the material dataset. This method has been extended to: identify the stress–strain relation of nonlinear elastic materials [13], problems with noisy material datasets [18], geometrically nonlinear problems [16], and dynamic problems [19]. Ibanez et al. [6, 7] proposed a different

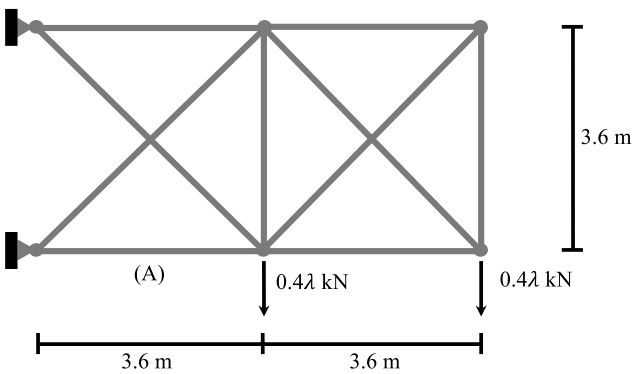

Fig. 5 10 bar truss example taken from [12]

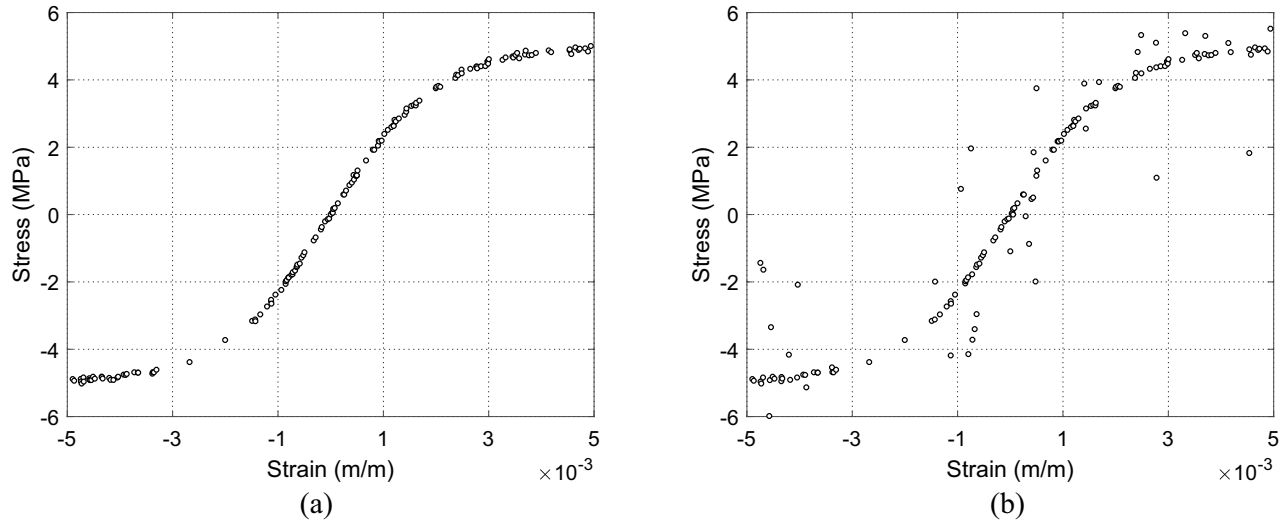

Fig. 6 Material datasets used for the numerical experiments. **a** Material dataset 1 which is not corrupted by outliers. **b** Material dataset 2 which is corrupted by outliers

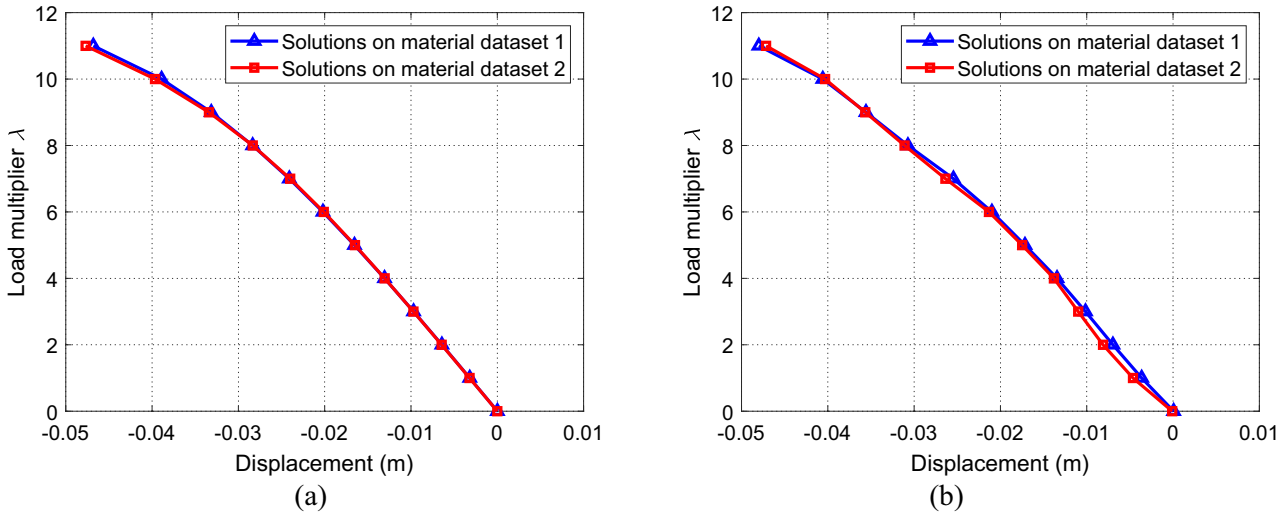

Fig. 7 Obtained equilibrium paths based on two material datasets. **a** Solutions of the proposed method. **b** Solutions obtained by the method in [12]

data-driven method, which utilized a manifold learning technique to extract the constitutive manifold from a material dataset. This method has been applied to thermodynamic consistency problems [5] and to correct the hyperelastic models from data [9]. A series of spline interpolation methods have been developed to identify the stored energy in hyper-elasticity based on experimental data [71–75]. The material dataset may not only be higly noisy but also is likely corrupted by outliers. Although some of those mentioned methods have been validated to be effective and robust in the context of noisy material datasets, it is unclear if they are still robust in the presence of outliers in the material dataset (note that noise is not necessarily representative of an outlier [25]). In this paper, since the results shown in Sects. 4.2 and 4.3 have demonstrated that the proposed approach has very good performance for data-driven regression in the presence of outliers, this section details the application of the proposed approach to data-driven computational mechanics with a nonlinear material dataset consisting of 150 data points and corrupted by outliers. Specifically, we test the properties of RLWLS-SVMR by way of analysis of truss structures. To achieve this, the RLWLS-SVMR is incorporated into the data-driven solver in [12]. Consider a truss structure with m bars and denote $\boldsymbol{u}$ and $\boldsymbol{p}$ as the nodal displacement vector and the external force vector, respectively.

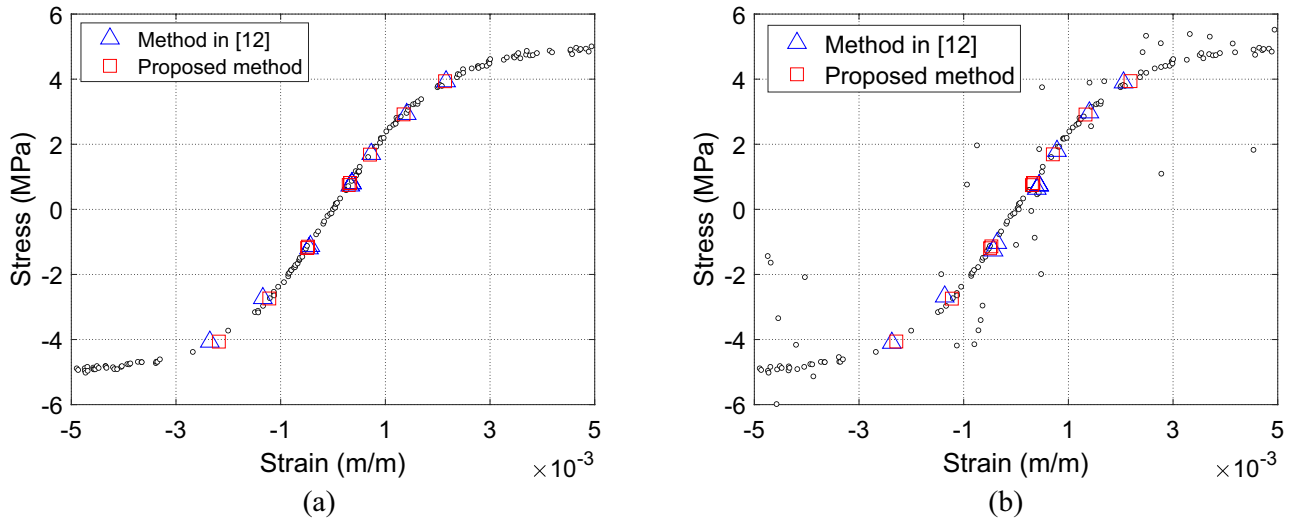

Fig. 8 Solutions obtained for $\lambda = 10$. **a** Solutions obtained based on Material dataset 1. **b** Solutions obtained based on Material dataset 2

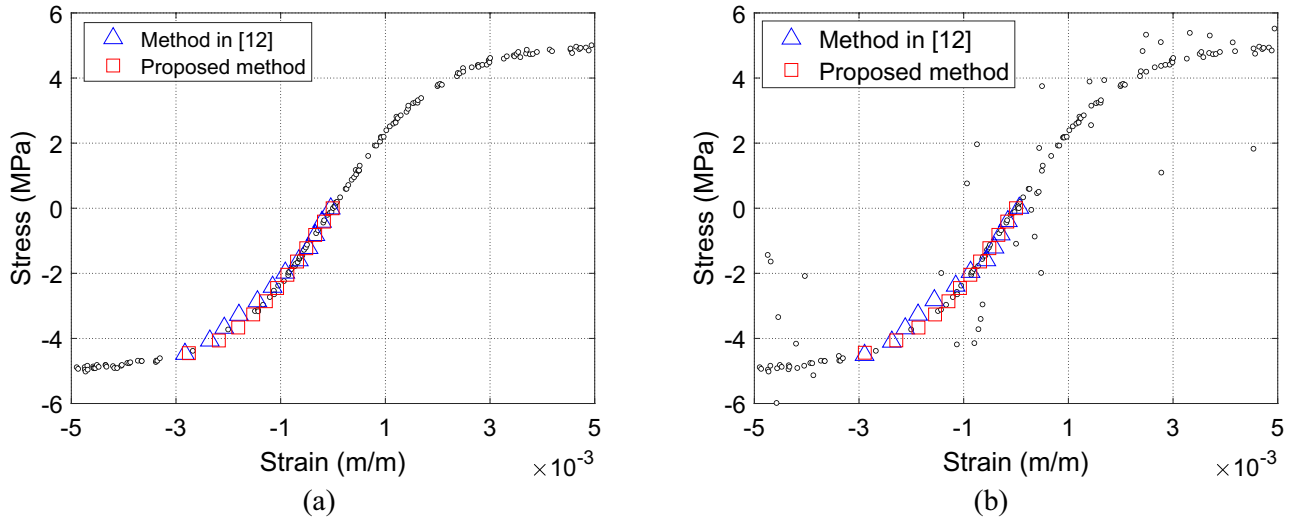

Fig. 9 Results comparison for obtained strains and stresses of member (A). **a** Results obtained based on material dataset 1. **b** Results obtained based on material dataset 2

The truss is subjected to compatibility conditions and force-balance equation constraints, which are given by

$$\varepsilon_i = \boldsymbol{b}_i^T \boldsymbol{u}, i = 1, \ldots, m \tag{14}$$

$$\sum_{i=1}^{m} v_i \sigma_i \boldsymbol{b}_i = \boldsymbol{p}, \tag{15}$$

where ε_i is the axial strain, σ_i is the axial stress, v_i is the volume of member i, and $\boldsymbol{b}_i$ $(i = 1, \ldots, m)$ are constant vectors.

For a given experimental material dataset, which is denoted as $\left\{\left(\check{\varepsilon}_t, \check{\sigma}_t\right)\right\}_{t=1}^{d}$, where $\check{\varepsilon}_t$ and $\check{\sigma}_t$ are observed uniaxial strain and stress values, respectively, and d is the number of observations. For each member $i = 1, \cdots, m$, given a material dataset $\left\{\left(\check{\varepsilon}_t, \check{\sigma}_t\right)\right\}_{t=1}^{d}$, the estimated stress $\hat{\sigma}_i$ at ε_i can be obtained using the proposed RLWLS-SVMR model, which is denoted as

$$\hat{\sigma}_i = f\left(\varepsilon_i; \left\{\left(\check{\varepsilon}_t, \check{\sigma}_t\right)\right\}_{t=1}^{d}\right), i = 1, \cdots, m. \tag{16}$$

Therefore, a data-driven solver for truss structures can be formulated by minimizing the following:

$$\text{Minimize} : \left\|\sigma - \hat{\sigma}\right\| \tag{17a}$$

$$\text{Subject to} : \text{Eqs.(14), (15), and (16).} \tag{17b}$$

According to [12], Eq. (17a) and [18] can be reduced, such that it only solves a set of nonlinear equations. We utilize this method to solve Eq. (17a) [18] and the detailed implementation procedure can be found in [12].

A 10 bar truss taken from [12] is used as an illustrative example for validating the proposed data-driven solver. As shown in Fig. 5, the 10 bar truss is comprised of members with cross-sectional area of 2000 mm^2. Two vertical external forces of 0.4λ kN are applied at the bottom nodes, where λ is a load multiplier. Figure 6 shows two material datasets where each is composed of 150 data points. Material dataset 1, shown in Fig. 6(a), is taken from [12] and not corrupted by outliers. As a contrast, Fig. 6(b) shows Material dataset 2, which is the same as Material dataset 1, but corrupted by outliers. The proposed data-driven solver is used to solve the 10 bar truss using these two material datasets. Additionally, the data-driven solver in [12] is used for comparison.

Figure 7 shows the obtained equilibrium paths, where the variation of the vertical displacement for the bottom rightmost node is presented. Figure 7(a) shows the equilibrium paths obtained by the proposed data-driven solver based on the two material datasets. It is observed in Fig. 7(a) that although Material dataset 2 is corrupted by outliers, the proposed data-driven solver can still be used to obtain the solutions, which are almost the same as the solutions obtained based on Material dataset 1. This result demonstrates that the presence of outliers does not alter the solutions obtained by proposed data-driven solver. Figure 7(b) presents the equlibrium paths obtained by the method in [12]. By observation, the data-driven solver in [12] is affected by outliers, which is apparent based on the discrepancy between the solutions obtained based on Material datasets 1 and 2. Figure 8 shows a comparison of the solutions obtained for $\lambda = 10$. From Fig. 8(a) and (b), it is evident that both methods display some robustness to outliers. Figure 9 depicts the solutions for member A, as shown in Fig. 5. By comparing the results in Fig. 9(a) with those in Fig. 9(b), it is observed that the proposed data-driven solver obtains nearly the same solutions for both material datasets, while the method in [12] obtains slightly different solutions. All of these comparisons illustrate that the proposed data-driven solver is robust against outliers in a material dataset and is also more robust than the method proposed in [12].

5 Conclusions

A novel robust ML approach is proposed for data-driven predictions in solving engineering problems, which is robust to input data corrupted by outliers. The proposed method is formulated as an optimization problem by coupling LWLS-SVMR with one weight function to overcome the LWLS-SVMR's drawback regarding lack of robustness to outliers close to query points, significantly reducing the negative interference of outliers. The formulation and implementation of the proposed method are introduced in detail. Furthermore, this method is a robust, local model, where prediction of a query point only requires the fitting of a subset (not the entire training dataset) where the data points are relevant to the query point. In comparison to other robust, global approaches, this characteristic enables avoidance of a potential negative influence from irrelevant points and achieves a suitable trade-off between the capacity of the learning system and the size of the training dataset. Four one-dimensional simulated datasets corrupted by non-extreme and extreme outliers and four multi-dimensional real-world engineering datasets are employed to verify that the proposed approach is able to significantly reduce the negative effects of outliers. The proposed RLWLS-SVMR exhibits robustness to outliers and performs best in comparison to the robust, global approaches in solving engineering problems. Furthermore, the proposed method is applied to produce a data-driven solver for structural analysis with a nonlinear material dataset corrupted by outliers. A truss structure is used to test the properties of the proposed data-driven solver. The results show that the proposed data-driven solver is robust against the presence of outliers in a material dataset.

Acknowledgements This material is based in part on work supported by the Natural Science Foundation of Hubei Province under Grant #2022CFB294, the National Natural Science Foundation of China under Grant #52208485 and the National Science Foundation under Grant CMMI #1944301. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Natural Science Foundation of Hubei Province, National Natural Science Foundation of China and National Science Foundation.

Data availability The data that support the findings of this study are openly available in UCI machine learning repository at https://archive.ics.uci.edu/ml/index.php.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Montáns FJ, Chinesta F, Gómez-Bombarelli R, Kutz JN (2019) Data-driven modeling and learning in science and engineering. Comptes Rendus Mécanique 347(11):845–855
2. Bock FE, Aydin RC, Cyron CJ, Huber N, Kalidindi SR, Klusemann B (2019) A review of the application of machine learning and data mining approaches in continuum materials mechanics. Front Mater 6:110

3. Adeli H (2001) Neural networks in civil engineering: 1989–2000. Comput-Aid Civil Infrast Eng 16(2):126–142
4. Reich Y (1997) Machine learning techniques for civil engineering problems. Comput-Aid Civil Infrast Eng 12(4):295–310
5. González D, Chinesta F, Cueto E (2019) Thermodynamically consistent data-driven computational mechanics. Contin Mech Thermodyn 31(1):239–253
6. Ibañez R, Borzacchiello D, Aguado JV, Abisset-Chavanne E, Cueto E, Ladevèze P, Chinesta F (2017) Data-driven non-linear elasticity: constitutive manifold construction and problem discretization. Comput Mech 60(5):813–826
7. Ibanez R, Abisset-Chavanne E, Aguado JV, Gonzalez D, Cueto E, Chinesta F (2018) A manifold learning approach to data-driven computational elasticity and inelasticity. Archiv Comput Methods Eng 25(1):47–57
8. González D, García-González A, Chinesta F, Cueto E (2020) A data-driven learning method for constitutive modeling: application to vascular hyperelastic soft tissues. Materials 13(10):2319
9. González D, Chinesta F, Cueto E (2019) Learning corrections for hyperelastic models from data. Front Mater 2019(6):14
10. Kanno Y (2020) A kernel method for learning constitutive relation in data-driven computational elasticity. Jpn J Ind Appl Math. https://doi.org/10.1007/s13160-020-00423-1
11. Kanno Y (2018) Simple heuristic for data-driven computational elasticity with material data involving noise and outliers: a local robust regression approach. Jpn J Ind Appl Math 35(3):1085–1101
12. Kanno Y (2018) Data-driven computing in elasticity via kernel regression. Theor Appl Mech Lett 8(6):361–365
13. Leygue A, Coret M, Réthoré J, Stainier L, Verron E (2018) Data-based derivation of material response. Comput Methods Appl Mech Eng 331:184–196
14. Versino D, Tonda A, Bronkhorst CA (2017) Data driven modeling of plastic deformation. Comput Methods Appl Mech Eng 318:981–1004
15. Capuano G, Rimoli JJ (2019) Smart finite elements: A novel machine learning application. Comput Methods Appl Mech Eng 345:363–381
16. Nguyen LTK, Keip MA (2018) A data-driven approach to nonlinear elasticity. Comput Struct 194:97–115
17. Kirchdoerfer T, Ortiz M (2016) Data-driven computational mechanics. Comput Methods Appl Mech Eng 304:81–101
18. Kirchdoerfer T, Ortiz M (2017) Data driven computing with noisy material data sets. Comput Methods Appl Mech Eng 326:622–641
19. Kirchdoerfer T, Ortiz M (2018) Data-driven computing in dynamics. Int J Numer Meth Eng 113(11):1697–1710
20. Gandomi AH, Roke DA (2015) Assessment of artificial neural network and genetic programming as predictive tools. Adv Eng Softw 88:63–72
21. Gandomi AH, Mohammadzadeh D, Pérez-Ordóñez JL, Alavi AH (2014) Linear genetic programming for shear strength prediction of reinforced concrete beams without stirrups. Appl Soft Comput 19:112–120
22. Nguyen H, Nguyen NM, Cao MT, Hoang ND, Tran XL (2021) Prediction of long-term deflections of reinforced-concrete members using a novel swarm optimized extreme gradient boosting machine. Eng Comput 38(2):1–13
23. Cheng MY, Gosno RA (2020) Symbiotic polyhedron operation tree (SPOT) for elastic modulus formulation of recycled aggregate concrete. Eng Comput 37(2):1–16
24. James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. springer, New York, p 18
25. Rousseeuw PJ, Leroy AM (1987) Robust regression and outlier detection. Wiley, New York
26. Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (2011) *Robust statistics: the approach based on influence functions.* John Wiley & Sons. 196.
27. Rousseeuw PJ (1984) Least median of squares regression. J Am Stat Assoc 79(388):871–880
28. Rousseeuw P, Yohai V (1984) Robust regression by means of S-estimators. Robust and nonlinear time series analysis. Springer, New York, pp 256–272
29. Rousseeuw PJ, Hubert M (2011) Robust statistics for outlier detection. Wiley Interdiscip Rev Data Min Knowled Dis 1(1):73–79
30. Mu HQ, Yuen KV (2015) Novel outlier-resistant extended Kalman filter for robust online structural identification. J Eng Mech 141(1):04014100
31. Yuen KV, Mu HQ (2012) A novel probabilistic method for robust parametric identification and outlier detection. Probab Eng Mech 30:48–59
32. Yuen KV, Ortiz GA (2017) Outlier detection and robust regression for correlated data. Comput Methods Appl Mech Eng 313:632–646
33. Rusiecki A (2007) Robust LTS backpropagation learning algorithm. International Work-Conference on Artificial Neural Networks. Springer, Berlin Heidelberg, pp 102–109
34. Roy MH, Larocque D (2012) Robustness of random forests for regression. J Nonparamet Statist 24(4):993–1006
35. Suykens J, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J (2002) Least Squares Support Vector Machines. World Scientific
36. Pal M, Deswal S (2011) Support vector regression based shear strength modelling of deep beams. Comput Struct 89(13–14):1430–1439
37. Pal M, Singh NK, Tiwari NK (2011) Support vector regression based modeling of pier scour using field data. Eng Appl Artif Intell 24(5):911–916
38. Nhu VH, Hoang ND, Duong VB, Vu HD, Bui DT (2020) A hybrid computational intelligence approach for predicting soil shear strength for urban housing construction: a case study at Vinhomes Imperia project, Hai Phong city (Vietnam). Eng Comput 36(2):603–616
39. Tran TH, Nguyen H, Nhat-Duc H (2019) A success history-based adaptive differential evolution optimized support vector regression for estimating plastic viscosity of fresh concrete. Eng Comput 37(2):1–14
40. Chou JS, Ngo NT, Pham AD (2015) Shear strength prediction in reinforced concrete deep beams using nature-inspired metaheuristic support vector regression. J Comput Civ Eng 30(1):04015002
41. Chou JS, Pham AD (2015) Smart artificial firefly colony algorithm-based support vector regression for enhanced forecasting in civil engineering. Comput-Aid Civil Infrastruct Eng 30(9):715–732
42. Prayogo D, Cheng MY, Wu YW, Tran DH (2020) Combining machine learning models via adaptive ensemble weighting for prediction of shear capacity of reinforced-concrete deep beams. Eng Comput 36(3):1135–1153
43. Cheng MY, Hoang ND (2012) Risk score inference for bridge maintenance project using evolutionary fuzzy least squares support vector machine. J Comput Civ Eng 28(3):04014003
44. Hoang ND, Nguyen QL (2019) A novel method for asphalt pavement crack classification based on image processing and machine learning. Eng Comput 35(2):487–498
45. Luo H, Paal SG (2018) Machine learning-based backbone curve model of reinforced concrete columns subjected to cyclic loading reversals. J Comput Civ Eng 32(5):04018042
46. Luo H, Paal SG (2019) A locally weighted machine learning model for generalized prediction of drift capacity in seismic vulnerability assessments. Comput-Aided Civil Infrastruct Eng. 34(11):1–16
47. Luo H, Paal SG (2021) Reducing the effect of sample bias for small data sets with double-weighted support vector transfer regression. Comput-Aided Civil Infrastruct Eng 36(3):248–263

48. Vapnik V (1995) The nature of statistical learning theory. Springer-Verlag, New York
49. Suykens JA, De Brabanter J, Lukas L, Vandewalle J (2002) Weighted least squares support vector machines: robustness and sparse approximation. Neurocomputing 48(1–4):85–105
50. De Brabanter K, Pelckmans K, De Brabanter J, Debruyne M, Suykens JA, Hubert M, De Moor B (2009) Robustness of kernel based regression: a comparison of iterative weighting schemes. International Conference on Artificial Neural Networks. Springer, Berlin. Heidelberg, pp 100–110
51. Menzies T, Butcher A, Marcus A, Zimmermann T, Cok D (2011) Local vs. global models for effort estimation and defect prediction. In: *2011 26th IEEE/ACM International Conference on Automated Software Engineering (ASE 2011)*. IEEE. 343–351
52. Hand DJ, Vinciotti V (2003) Local versus global models for classification problems: fitting models where it matters. Am Stat 57(2):124–131
53. Bottou L, Vapnik V (1992) Local learning algorithms. Neural Comput 4(6):888–900
54. Vapnik V, Bottou L (1993) Local algorithms for pattern recognition and dependencies estimation. Neural Comput 5(6):893–909
55. Karevan Z, Feng Y, Suykens JA (2017) Moving Least Squares Support Vector Machines for weather temperature prediction. In: *Proc. of the European Symposium on Artificial Neural Networks (ESANN). Bruges, Belgium.*
56. Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74(368):829–836
57. Cleveland WS, Devlin SJ (1988) Locally weighted regression: an approach to regression analysis by local fitting. J Am Stat Assoc 83(403):596–610
58. Atkeson CG, Moore AW, Schaal S (1997) Locally weighted learning. Artif Intell Rev 11:11–73
59. Atkeson CG, Moore AW, Schaal S (1997) Locally weighted learning for control. Artif Intell Rev 11(1–5):75–113
60. Suykens JAK, Lukas L, Van Dooren P, De Moor B, Vandewalle J (1999) Least squares support vector machine classifiers: a large scale algorithm. In: *European Conference on Circuit Theory and Design, ECCTD*. Citeseer. 839–842
61. De Brabanter J, Pelckmans K, Suykens JA, Vandewalle J (2002) Robust cross-validation score function for non-linear function estimation. In *International Conference on Artificial Neural Networks*. 713-719. Springer. Berlin, Heidelberg
62. Kvålseth TO (1985) Cautionary note about R^2. Am Stat 39(4):279–285
63. Liu J, Wang Y, Fu C, Guo J, Yu Q (2016) A robust regression based on weighted LSSVM and penalized trimmed squares. Chaos, Solitons Fractals 89:328–334
64. Yang X, Tan L, He L (2014) A robust least squares support vector machine for regression and classification with noise. Neurocomputing 140:41–52
65. Huber PJ (1964) Robust estimation of a location parameter. Ann Mathemat Statis. 35(1):73–101
66. Berry M, Parrish M, Eberhard M (2004) PEER Structural Performance Database. University of California, Berkeley, User's Manual
67. Quinlan JR (1993). Combining instance-based and model-based learning. In: *Proceedings of the tenth international conference on machine learning*. 236–243.
68. Nelson W (1981) Analysis of performance-degradation data from accelerated tests. IEEE Trans Reliab 30(2):149–155
69. Stone WC, Cheok GS (1989) *Inelastic behavior of full-scale bridge columns subjected to cyclic loading, NIST BSS 166*. U.S.National Institute of Standards and Technology, Gaithersburg, MD, p 261
70. Priestley MJN, Potangaroa RT, Park R (1981) Ductility of spirally-confined concrete columns. J Struct Div 107(1):181–202
71. Sussman T, Bathe KJ (2009) A model of incompressible isotropic hyperelastic material behavior using spline interpolations of tension–compression test data. Commun Numer Methods Eng 25(1):53–63
72. Latorre M, Montáns FJ (2013) Extension of the Sussman-Bathe spline-based hyperelastic model to incompressible transversely isotropic materials. Comput Struct 122:13–26
73. Crespo J, Latorre M, Montáns FJ (2017) WYPIWYG hyperelasticity for isotropic, compressible materials. Comput Mech 59(1):73–92
74. De Rosa E, Latorre M, Montáns FJ (2017) Capturing anisotropic constitutive models with WYPiWYG hyperelasticity; and on consistency with the infinitesimal theory at all deformation levels. Int J Non-Linear Mech 96:75–92
75. Latorre M, Montáns FJ (2020) Experimental data reduction for hyperelasticity. Comput Struct 232:105919
76. Bai JL, He J, Li C, Jin SS, Yang H (2022) Experimental investigation on the seismic performance of a novel damage-control replaceable RC beam-to-column joint. Engineering Structures https://doi.org/10.1016/j.engstruct.2022.114692
77. Zhou Y, Chen LZ, Long L (2023) Modeling cyclic behavior of squat reinforced concrete walls exposed to acid deposition. Journal of Building Engineering https://doi.org/10.1016/j.jobe.2022.105432