

Journal of Materials and Engineering Structures

Research Paper

Comparative analysis of Isolation Forest and One-Class SVM for outlier detection in GNSS-RTK cable-stayed bridge monitoring data

Khanh Giang Le [a,*]

[a] *Faculty of Civil Engineering, University of Transport and Communications, Hanoi, Vietnam*

ARTICLE INFO

Article history :

Received : 22 November 2025

Revised : 24 February 2026

Accepted : 28 February 2026

Keywords:

GNSS-RTK, Outliers

Isolation Forest

One-Class SVM

Displacement Monitoring

Cable-Stayed Bridges

ABSTRACT

Accurate GNSS-RTK displacement data are essential for reliable structural health monitoring (SHM) of long-span cable-stayed bridges; however, these measurements are often corrupted by noise, multipath, and gross errors. This study conducts a comparative evaluation of two unsupervised anomaly detection algorithms—Isolation Forest (IF) and One-Class Support Vector Machine (OC-SVM)—using a dataset of 2,078 GNSS-RTK displacement samples collected at the midpoint of the Can Tho cable-stayed bridge. The results show that both algorithms successfully identified major abnormal measurements with 70–73% overlap in detected outliers across the X, Y, and Z axes. IF detected 101–104 outliers, while OC-SVM detected 98–105, depending on the axis. IF consistently achieved higher cluster separability, with Silhouette Scores of 0.678 (X), 0.571 (Y), and 0.453 (Z), compared with 0.672, 0.563, and 0.433 for OC-SVM. Moreover, IF significantly outperformed OC-SVM in reconstruction fidelity, yielding Reconstruction Errors of 0.600–0.839, markedly lower than the 3.060–3.914 obtained by OC-SVM. After filtering out the detected anomalies, the GNSS-RTK displacement series showed substantial improvements in data stability; the standard deviation was reduced by 44.24% (X), 18.90% (Y), and 12.08% (Z) for IF and by 43.98% (X), 17.98% (Y), and 10.51% (Z) for OC-SVM. These quantitative findings demonstrate that both methods enhance GNSS-RTK data quality, while the Isolation Forest algorithm offers superior robustness, better cluster discrimination, and improved preservation of true structural responses. The results confirm the suitability of IF as a reliable, computationally efficient preprocessing tool for real-time GNSS-based SHM of cable-stayed bridges.

1 Introduction

Cable-stayed bridges are critical components of modern transportation infrastructure, requiring continuous and accurate structural health monitoring (SHM) to ensure their safety, functionality, and longevity [1]. Among various monitoring technologies, the Global Navigation Satellite System–Real-Time Kinematic (GNSS-RTK) technique has emerged as a valuable tool for high-precision displacement measurements of large-scale bridge structures [2]. GNSS observations have been widely used for displacement monitoring of cable-stayed bridges in Vietnam, providing high-precision time-series data

* *Corresponding author. Tel.: +84 983663031.*
E-mail address: gianglk@utc.edu.vn

e-ISSN: 2170-127X, (cc) BY-SA

RESEARCH REVIEW of Sciences and Technologies

for structural health analysis [3]. Previous studies have confirmed the capability of high-rate GNSS for dynamic monitoring. Moschas and Stiros (2013) demonstrated that high-frequency GPS observations are largely free from colored noise and suitable for capturing dynamic signals [4], while Moschas and Stiros (2015) further showed that 100-Hz GNSS data can resolve millimeter-level bridge deflections and vibration characteristics [5]. However, GNSS-RTK data are often affected by noise, multipath effects, and gross errors, which can lead to inaccurate assessments and potential misinterpretations of bridge behavior [6].

Outlier detection has therefore become a crucial preprocessing step in GNSS-RTK data analysis to enhance data quality and reliability [7]. Traditional filtering and threshold-based statistical methods often fail to capture the complex, non-linear, and non-stationary characteristics of GNSS-RTK time series, especially under varying environmental conditions [8]. In recent years, machine learning-based approaches have gained increasing attention due to their ability to model nonlinear behaviors and adapt to diverse noise characteristics [9]. Both supervised and unsupervised techniques provide powerful tools for identifying abnormal measurements and improving the accuracy of displacement monitoring systems [10]. For example, Ho et al. (2025) proposed an unsupervised learning framework combining t-SNE and adaptive dynamic time warping barycenter averaging (ADBA) to analyze GNSS–RTK displacement data from the Tran Thi Ly cable-stayed bridge in Vietnam, demonstrating improved cluster separability and interpretable detection of structural behavior patterns without labeled data [11]. Similarly, Tran et al. (2025) analyzed vertical GNSS-RTK displacement time series from the Nhat Tan cable-stayed bridge in Vietnam, using Hierarchical Agglomerative Clustering (HAC) and Gaussian Mixture Model (GMM) clustering, identifying two dominant structural displacement states and showing that GMM achieved superior cluster separability [12]. These works confirm the feasibility of using unsupervised learning for early anomaly detection and proactive bridge maintenance.

Globally, several studies have explored the use of data-driven and machine learning techniques for GNSS-based or vibration-based structural monitoring [30]. Shen et al. (2023) developed an enhanced K-means clustering approach to detect three-dimensional displacements from GNSS kinematic positioning data in deformation monitoring experiments. Their simulation and field tests showed that the proposed method enables millimeter-level accuracy in extracting 3D displacements from GNSS-RTK observations [13]. Other researchers compared multiple unsupervised learning algorithms, including K-nearest neighbors, Gaussian mixture models, One-Class Support Vector Machines (OC-SVM), density-based clustering, and deep autoencoders, revealing that OC-SVM is a reliable method for detecting structural damage such as loose bolts in steel bridge models [9]. Long and Buyukozturk (2014) developed an automated OC-SVM-based damage detection approach and validated it using vibration measurements from a steel frame laboratory structure, showing that machine-learning-based pattern recognition can effectively identify structural damage from sensor data [14].

For cable-stayed bridges, Çetindemir et al. (2022) applied OC-SVM and Local Outlier Factor (LOF) to statistical indicators extracted from long-term acceleration data of the Osman Gazi bridge in Turkey, demonstrating that anomaly detection can be achieved efficiently without processing raw vibration signals [15]. Nguyen et al. (2014) applied dimensionality reduction and SVM-based pattern recognition to vibration data from a laboratory structure and the Sydney Harbour Bridge, showing that random projection can accelerate damage detection while maintaining accuracy in SHM applications [16]. Furthermore, Bayane et al. (2024) proposed an unsupervised machine-learning framework for real-time bridge damage detection using strain and acceleration monitoring data. The approach was validated on a steel bascule railway bridge in Sweden that experienced an actual brittle cracking event, showing that anomaly-detection algorithms such as Isolation Forest (IF), OC-SVM, LOF, can reliably identify abrupt structural changes from continuous monitoring data. Among these methods, IF demonstrated the best performance in detecting structural anomalies [17].

Other studies have also demonstrated the strong capability of IF in infrastructure monitoring. Sun et al. (2022) employed this algorithm to detect malfunctioning components such as bearings, expansion joints, and dampers in the Jiangyin bridge in China based on cumulative displacement anomalies [10]. Similarly, Mousavi and Gandomi (2022) combined Variational Mode Decomposition and IF for unsupervised bridge monitoring and validated the method on the Z24 benchmark, showing that damage can be detected without baseline data [18]. These findings indicate that unsupervised learning algorithms, particularly IF and OC-SVM, have shown great potential in structural anomaly detection. Nevertheless, their application to GNSS-RTK displacement data for outlier removal in cable-stayed bridge monitoring remains largely unexplored.

In Vietnam, research on GNSS-RTK outlier detection is still in its early stages. In a case study on GNSS monitoring data from the Can Tho cable-stayed bridge in Vietnam, Le and Vo (2023) applied the Hampel identifier to detect and correct abnormal displacement values in the time series [19]**.** In a landslide monitoring study in Vietnam, Duong et al. (2025)

combined GNSS-RTK observations with post-processing and Extended Kalman filtering to improve the accuracy and stability of displacement monitoring results [20]. For continuous GNSS coordinate time series, Tran et al. (2016) investigated several outlier detection approaches, including geometric window-based methods, regression modeling, and Kalman filtering, and found that Kalman filtering generally provided the most reliable results for tectonic deformation studies [21]. In a machine-learning-based outlier detection study, Nguyen and Tran (2023) tested IF, OC-SVM, and LOF on a simulated GNSS position time series and demonstrated that IF achieved the best performance in detecting outliers [22]. Despite these advancements, no existing research has systematically compared IF and OC-SVM for GNSS-RTK outlier detection in cable-stayed bridge monitoring in Vietnam.

To fill this research gap, this study conducts a comparative analysis of two state-of-the-art unsupervised learning algorithms - Isolation Forest and One-Class Support Vector Machine -for outlier detection in GNSS-RTK displacement time series obtained from a cable-stayed bridge in Vietnam. The specific objectives are to:

1. Develop a GNSS-RTK outlier detection framework based on IF and OC-SVM;
2. Evaluate and compare their detection performance using field-measured GNSS-RTK data; and
3. Assess the impact of outlier removal on the accuracy and reliability of GNSS-RTK-based displacement monitoring.

This study contributes to the growing body of knowledge on data-driven SHM by introducing a practical comparative framework tailored to GNSS-RTK monitoring systems for cable-stayed bridges in Vietnam. The proposed approach enhances the quality of GNSS-RTK measurements and provides insights for future integration of machine learning techniques into real-time bridge monitoring and maintenance decision-making.

2 Research data and methods

2.1 Research data

2.1.1 Bridge description and data collection

The Can Tho cable-stayed bridge, located in southern Vietnam, serves as the case study for this research. The bridge has a total width of 26 m, a main span of 550 m, and a total length of 2,750 m. The height of the tower measured from the bridge deck is 134.7 meters, and the bridge was designed according to the AASHTO LRFD Bridge Design Specifications [31]. The central 210 m section of the main span consists of prefabricated steel box girders connected to cast-in-place concrete girders at both ends.

In 2010, the bridge was equipped with an integrated Structural Health Monitoring (SHM) system comprising accelerometers, strain gauges, anemometers, and thermometers to record structural and environmental responses in real time. The GNSS-RTK subsystem includes two base stations and nine rover receivers, primarily installed along the main span. The GNSS monitoring operates continuously in RTK mode, providing 1 Hz sampling frequency data. The measurements are automatically averaged and stored at intervals of 1 minute, 10 minutes, 1 hour, and 1 day on a central control computer. According to system specifications, the measurement accuracy is approximately ±(10 mm + 1 ppm) in the horizontal direction and ±(20 mm + 1 ppm) in the vertical direction.

Previous studies on the Can Tho Bridge [11] identified four critical monitoring locations that represent the overall deformation behavior of the structure: the tops of the two pylons, the mid-point of the main span, and the quarter-span point. In this study, GNSS-RTK data from these four monitoring positions were collected and analyzed, as illustrated in Figure 1, which shows the placement of GNSS antennas, anemometers, and air temperature sensors across the bridge deck.

For the outlier detection experiment, GNSS-RTK displacement data at the midpoint of the main span were used, covering the X (longitudinal), Y (transverse), and Z (vertical) directions. The dataset consists of 2,078 ten-minute averaged observations collected continuously from 01 January to 18 February 2016, representing normal operating conditions of the bridge. While most of the measurements remained stable, several samples exhibited abnormal fluctuations due to temporary signal interruptions or environmental interference. The GNSS-RTK coordinates are expressed in the bridge's local coordinate system, where the X-axis aligns with the bridge's longitudinal direction, Y with the transverse direction, and Z with the vertical displacement.

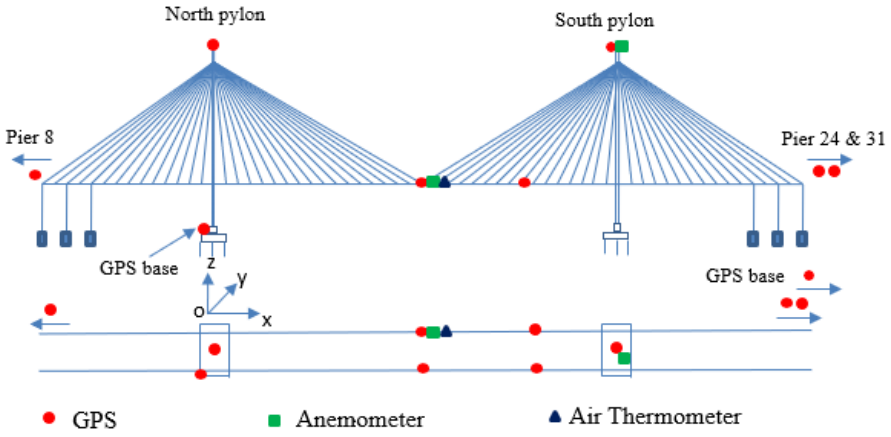


Fig. 1 - Layout of GNSS receivers on the Can Tho cable-stayed bridge.

2.1.2 Data preprocessing

Before applying machine learning algorithms, the GNSS-RTK displacement data were preprocessed to ensure quality, consistency, and suitability for outlier detection. Raw GNSS-RTK observations may contain missing or invalid samples due to temporary satellite signal loss, multipath effects, or atmospheric interference. In this study, short data interruptions of fewer than 20 minutes were linearly interpolated to maintain time-series continuity, following best practices in GNSS-based bridge monitoring [13]. This threshold, equivalent to approximately two 10-minute sampling intervals, was sufficient to preserve deformation trends without distorting the underlying displacement signal. Longer gaps were excluded from further analysis to avoid interpolation bias or artificial drift. All datasets were visually inspected to ensure temporal alignment and completeness prior to subsequent processing. After applying this procedure, inspection results confirmed that no data gaps exceeding 20 minutes remained in any of the GNSS-RTK displacement time series, indicating that the dataset was temporally continuous and suitable for subsequent analysis. The final dataset comprised 2,078 complete samples, ready for anomaly detection analysis.

2.2 Research methods

This study employed two unsupervised machine learning algorithms - Isolation Forest (IF) and One-Class Support Vector Machine (OC-SVM) - to detect outliers in GNSS-RTK displacement time series. Both algorithms are data-driven and require no prior labeling, making them suitable for GNSS-based Structural Health Monitoring where annotated data are typically unavailable.

2.2.1 Isolation Forest algorithm

The Isolation Forest (IF) is an ensemble-based unsupervised learning algorithm specifically designed for anomaly detection [23]. Unlike traditional approaches that model normal data distribution, IF isolates outliers based on the principle that anomalies are few and significantly different from the majority of observations.

The algorithm constructs multiple binary decision trees (isolation trees). At each node, a random feature and a random split value between that feature's minimum and maximum are selected. The number of splits required to isolate a data point serves as a measure of its abnormality - outliers typically have shorter average path lengths within the trees [24].

The Isolation Forest algorithm can be summarized as follows [24]:

(1) Randomly select a subsample from the dataset and determine the upper and lower bounds of each variable.

(2) Randomly choose a split value within the range of the selected feature.

(3) Divide the data into two subsets based on the split value.

(4) Repeat the process recursively until each leaf node contains only one data point or the maximum tree depth is reached.

After constructing the forest, the algorithm computes an anomaly score for each observation based on its average path length across all trees. The anomaly score $s(x,n)$ for a sample x from a dataset of size n is defined as [23]:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{1}$$

where:

$s(x,n)$ is monotonic with respect to $h(x)$;

(n) is the average path length of unsuccessful searches in a Binary Search Tree of n samples, used for normalization;

$(h(x))$ is the average path length of x across all isolation trees.

Interpretation of $s(x,n)$:

when $E(h(x)) \to 0$, $s(x,n) \to 1$: highly likely to be an anomaly.

when $E(h(x)) \to c(n)$, $s(x,n) \to 0.5$: uncertain or borderline.

when $E\big(h(x)\big) \to n-1$, $s(x,n) \to 0$: highly normal.

Using the anomaly score s, the following assessments can be made:

(a) Instances with s values very close to 1 are definitive anomalies.

(b) Instances with s values significantly less than 0.5 are likely to be normal.

(c) If all instances return s ≈ 0.5, the dataset likely lacks any distinct anomalies.

The Isolation Forest algorithm is highly efficient and effective for detecting outliers in large-scale or high-dimensional datasets, including GNSS coordinate time series used for structural health monitoring.

2.2.2 One-class Support Vector Machines Algorithm

The One-Class Support Vector Machine (OC-SVM) is a semi-supervised learning algorithm widely applied in anomaly and novelty detection tasks [25]. Unlike conventional Support Vector Machines (SVMs) used for binary or multiclass classification, OC-SVMs are trained only on data representing the normal class. Their objective is to learn a decision boundary that encloses the majority of the training samples in the feature space, thereby distinguishing normal data from potential outliers or novel observations [26].

The OC-SVM algorithm operates under the principle of maximizing the margin between the origin and the learned decision boundary in a high-dimensional feature space. It attempts to find a hyperplane that best separates the normal data points from the origin, effectively defining a small region that captures the "normal" behavior of the dataset while excluding abnormal points. This is achieved by solving the following optimization problem [27]:

$$min\frac{1}{2}\|w\|^2 + \frac{1}{vn}\sum_{i=1}^{n}\xi_i - \rho \tag{2}$$

subject to:

$$(w.\phi(x_i)) \geq \rho - \xi_i, \qquad \xi_i \geq 0 \tag{3}$$

where:

w is the weight vector normal to the separating hyperplane;

v (nu-parameter) controls the trade-off between the fraction of outliers and the size of the decision region;

n is number of data points;

ξ_i are slack variables allowing certain violations for noisy data;

ρ defines the offset from the origin; and

$\phi(x_i)$ denotes the kernel-induced feature mapping of input data x_i;

To handle nonlinear separability in the input space, OC-SVM employs kernel functions that implicitly map data into a higher-dimensional space. The Radial Basis Function (RBF) kernel is one of the most used due to its flexibility and robustness, defined as:

$$K\big(x_i, x_j\big) = \exp(-\gamma\big\|x_i - x_j\big\|^2) \tag{4}$$

where:

$K(x_i, x_j)$ is the kernel value between two points x_i and x_j.

$\|x_i - x_j\|^2$ is the squared Euclidean distance between the two points x_i and x_j.

γ is a kernel hyperparameter that determines the influence of a single training example:

A small γ value implies a broad influence, leading to smoother decision boundaries;

A large γ value yields a narrow influence, making the model more sensitive to local variations.

After training, the OC-SVM determines whether a new data point x belongs to the normal class or is an outlier by evaluating its position relative to the learned decision boundary. Points lying inside the boundary are considered normal, while those outside are classified as anomalies.

This method is particularly suitable for structural health monitoring scenarios, where only normal-state GNSS-RTK data are available during training. The OC-SVM effectively identifies subtle deviations or irregular displacement patterns that may indicate abnormal structural behavior.

2.2.3 Parameter settings

Table 1 summarizes the hyperparameters adopted for the Isolation Forest (IF) and One-Class SVM (OC-SVM) models. Parameter selection was guided by grid search and sensitivity analysis to ensure model stability and generalization. The contamination rate (0.05) and anomaly fraction (ν = 0.05) were chosen to reflect the expected proportion of outliers in the GNSS-RTK displacement dataset.

Table 1. Hyperparameter settings for Isolation Forest and One-Class SVM

Algorithm	Parameter	Value	Description
Isolation Forest	n_estimators	100	Number of base estimators (trees).
	contamination	0.05	Expected proportion of anomalies.
	random_state	42	Ensures reproducibility.
One-Class SVM	kernel	RBF	Captures nonlinear relationships.
	gamma	0.1	Kernel coefficient.
	nu	0.05	Upper bound on the fraction of anomalies.
	tol	1e−3	Convergence tolerance.
	max_iter	10,000	Maximum iterations.

Algorithm 1. Outlier Detection Workflow using Isolation Forest and One-Class SVM

Input : Preprocessed GNSS-RTK displacement dataset $D = \{D_x, D_y, D_z\}$

Output : Detected outlier sets O_{IF}, O_{SVM}, and overlapping outliers $O_{overlap}$

1 For each displacement axis $i \in \{x, y, z\}$:
Train Isolation Forest (IF) model with contamination rate = 0.05 and random state = 42.
Train One-Class SVM (OC-SVM) with kernel = RBF, $\gamma = 0.1$, and $\nu = 0.05$.
Predict outlier labels for each method using fit_predict().
Identify overlapping outliers: $O_{overlap} = O_{IF} \cap O_{SVM}$.

2 Compute evaluation metrics: Silhouette Score, Outlier Ratio, and Reconstruction Error for each axis.

3 Visualize and compare anomaly detections by IF, OC-SVM, and their intersection.

4 Save combined outlier indices and cleaned GNSS-RTK dataset for further structural analysis.

End Algorithm

All algorithms were implemented in Python (scikit-learn 1.4.1) using 10-minute averaged GNSS-RTK displacement data. Parameter tuning was performed through grid search to optimize Silhouette Score and minimize Reconstruction Error.

Algorithm 2. Data Quality Evaluation after Outlier Filtering

Input : Original GNSS-RTK dataset $D_{raw} = \{x, y, z\}$; filtered datasets D_{IF}, D_{SVM}
Output : Table of standard deviations and percentage improvements for each axis.

1	Import the raw and filtered datasets (Isolation Forest and One-Class SVM results).
2	For each displacement axis $i \in \{x, y, z\}$:
	(a) Compute the standard deviation before filtering σ_{before}.
	(b) Compute the standard deviation after filtering σ_{after}.
	(c) Calculate percentage improvement:
	$\mathrm{I} = 100 \times (\sigma_{before} - \sigma_{after})/\sigma_{before}$
3	Combine results for both algorithms and export to a CSV file for reporting.

End Algorithm

2.2.4 Evaluating the Effectiveness of Models

The performance of the outlier detection algorithms was evaluated using three complementary quantitative indicators: Silhouette Score (SS), Outlier Ratio (OR), and Reconstruction Error (RE). These metrics jointly assess clustering quality, detection sensitivity, and reconstruction fidelity, thereby providing a comprehensive evaluation of each model's effectiveness.

2.2.4.1 Silhouette Score (SS)

The Silhouette Score measures the compactness and separation of detected normal and outlier samples [28]. It is defined as [29]:

$$SS = \frac{b_i - a_i}{max(a_i, b_i)} \tag{5}$$

where: SS is Silhouette coefficients for i^{th} object; a_i is average of the minimum distance between i^{th} object in the same cluster; b_i is average of the minimum distance between i^{th} object in a different cluster.

The score ranges from −1 to 1, where higher values indicate better separation between normal and abnormal data.

In this study, a higher Silhouette Score represents a clearer distinction between GNSS-RTK displacement anomalies and typical structural responses.

2.2.4.2 Outlier Ratio (OR)

The Outlier Ratio quantifies the proportion of samples classified as outliers relative to the total number of observations [23]:

$$OR = \frac{N_{outlier}}{N_{total}} \times 100\% \tag{6}$$

A moderate outlier ratio indicates a balanced detection threshold, avoiding both excessive false alarms and missed anomalies.

Comparing OR between Isolation Forest and One-Class SVM helps assess model sensitivity and consistency across the three displacement components (X, Y, and Z).

2.2.4.3 Reconstruction Error (RE)

To assess how accurately each model represents the underlying structural behavior, the mean absolute reconstruction error between observed and predicted displacement representations was computed as:

$$\mathrm{RE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{x}_i| \tag{7}$$

where x_i is the observed displacement and $\hat{x}_i$ is the model-estimated decision function value.

Lower RE values indicate that the model better preserves the original displacement trend while effectively isolating anomalous measurements.

2.2.4.4 Comparative Evaluation

All three indicators were computed separately for each coordinate axis (X, Y, and Z). The Isolation Forest and One-Class SVM models were then compared in terms of:

Detection quality (via Silhouette Score),

Sensitivity and robustness (via Outlier Ratio), and

Model fidelity (via Reconstruction Error).

This combined evaluation framework provides a balanced and interpretable measure of outlier detection performance, ensuring that the selected algorithm not only identifies anomalies effectively but also maintains the structural integrity of the GNSS-RTK displacement data.

3 Results and discussions

3.1 Outlier Detection Results

The Isolation Forest (IF) and One-Class Support Vector Machine (OC-SVM) were applied to the preprocessed GNSS-RTK displacement data recorded at the midpoint of the main span of the Cần Thơ cable-stayed bridge. The dataset comprises 2,078 ten-minute averaged samples collected between 1 January and 18 February 2016. Analyses were conducted separately for the three displacement components (X: longitudinal, Y: transverse, Z: vertical) to assess algorithm sensitivity and robustness across motion directions.

3.1.1 X-axis (longitudinal) results

Figure 2 illustrates the comparative performance of the two unsupervised learning algorithms in identifying anomalous points along the longitudinal (X-axis) displacement component.

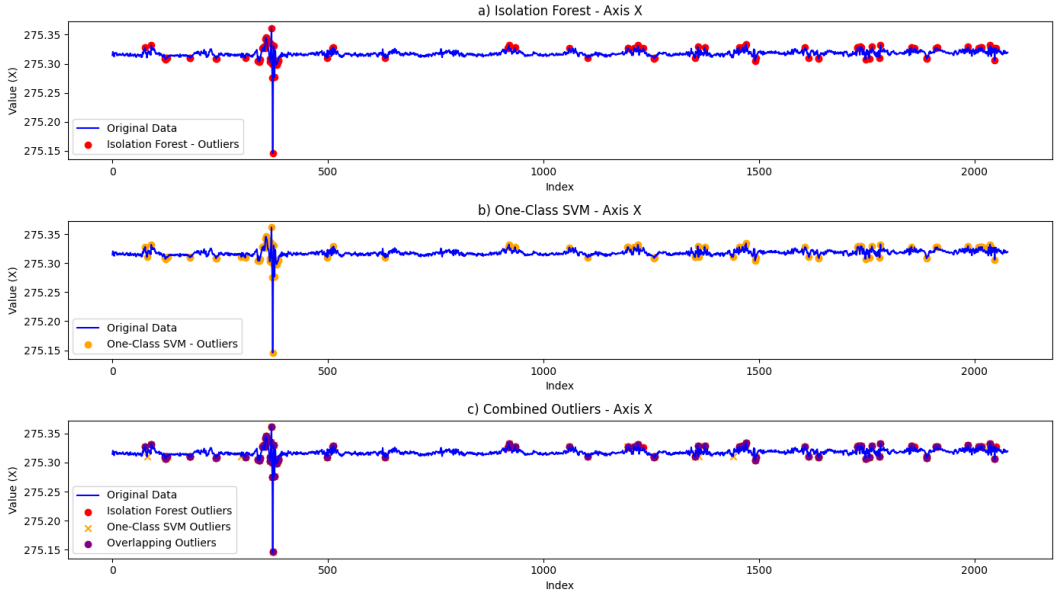


Fig. 2 - Outlier detection results for the X-axis GNSS-RTK displacement time series at the midpoint of the Can Tho cable-stayed bridge using (a) IF, (b) OC-SVM, and (c) the combined detection outcome.

As shown in Figure 2, both IF and OC-SVM reliably detected the principal anomalies in the longitudinal series. The most prominent abnormal cluster is observed near indices ≈ 400–450 and corresponds to a sudden displacement drop that both algorithms consistently flagged. IF (Figure 2a) produced well-localized detections, isolating abrupt spikes while remaining robust to small oscillations. OC-SVM (Figure 2b) also identified the dominant anomalies but flagged additional low-amplitude deviations, reflecting its higher sensitivity to local variations. In summary, IF offers a more selective detection pattern on the X-axis, whereas OC-SVM is more sensitive but prone to denser detections of minor fluctuations.

3.1.2 Y-axis (transverse) results

Figure 3 displays the transverse component results. Both algorithms captured the dominant anomaly cluster around indices ≈ 400–450, corresponding to a sharp transient in the recorded signal. IF (Figure 3a) detected a moderate number of clearly defined outliers concentrated at peaks and troughs where deviations from the local trend. OC-SVM (Figure 3b) produced similar primary detections but additionally marked a greater number of smaller oscillations as anomalies. The agreement on major events between IF and OC-SVM on the Y-axis corroborates their consistency for transverse displacement monitoring.

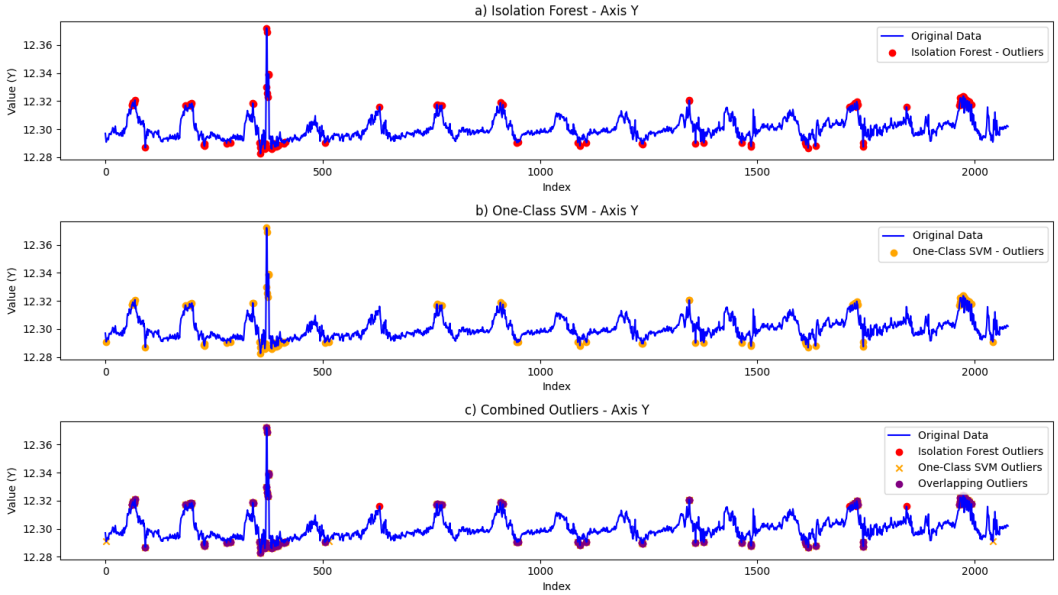


Fig. 3 - Outlier detection results for the Y-axis GNSS-RTK displacement time series at the midpoint of the Can Tho cable-stayed bridge using (a) IF, (b) OC-SVM, and (c) the combined detection outcome.

3.1.3 Z-axis (vertical) results

The vertical component (Figure 4) exhibits a pronounced diurnal pattern associated with thermal effects, superimposed with intermittent abrupt deviations likely caused by multipath or short-term reception issues.

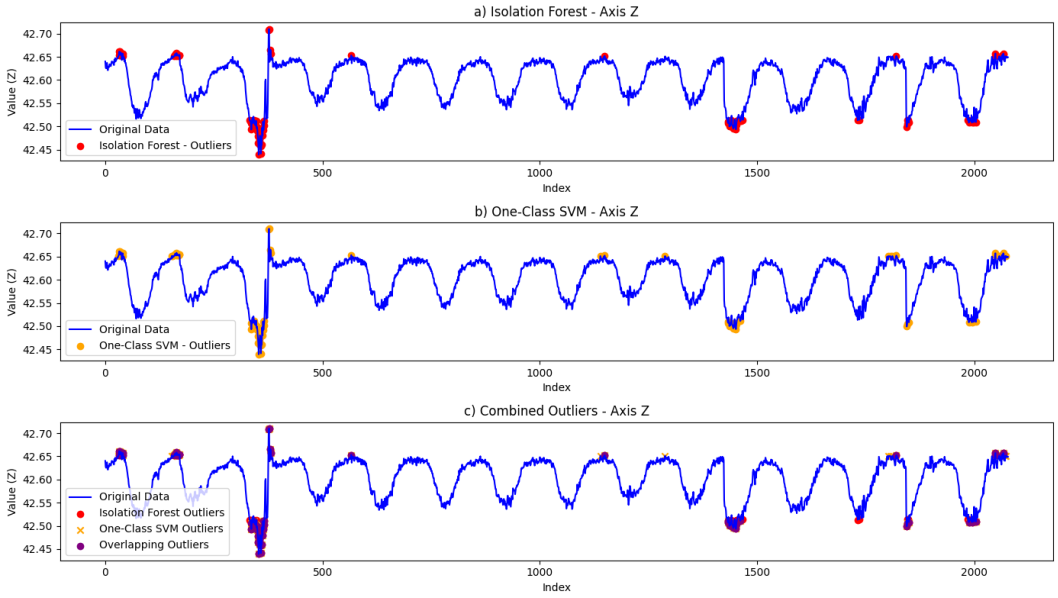


Fig. 4 - Outlier detection results for the Z-axis GNSS-RTK displacement time series at the midpoint of the Can Tho cable-stayed bridge using (a) IF, (b) OC-SVM, and (c) the combined detection outcome.

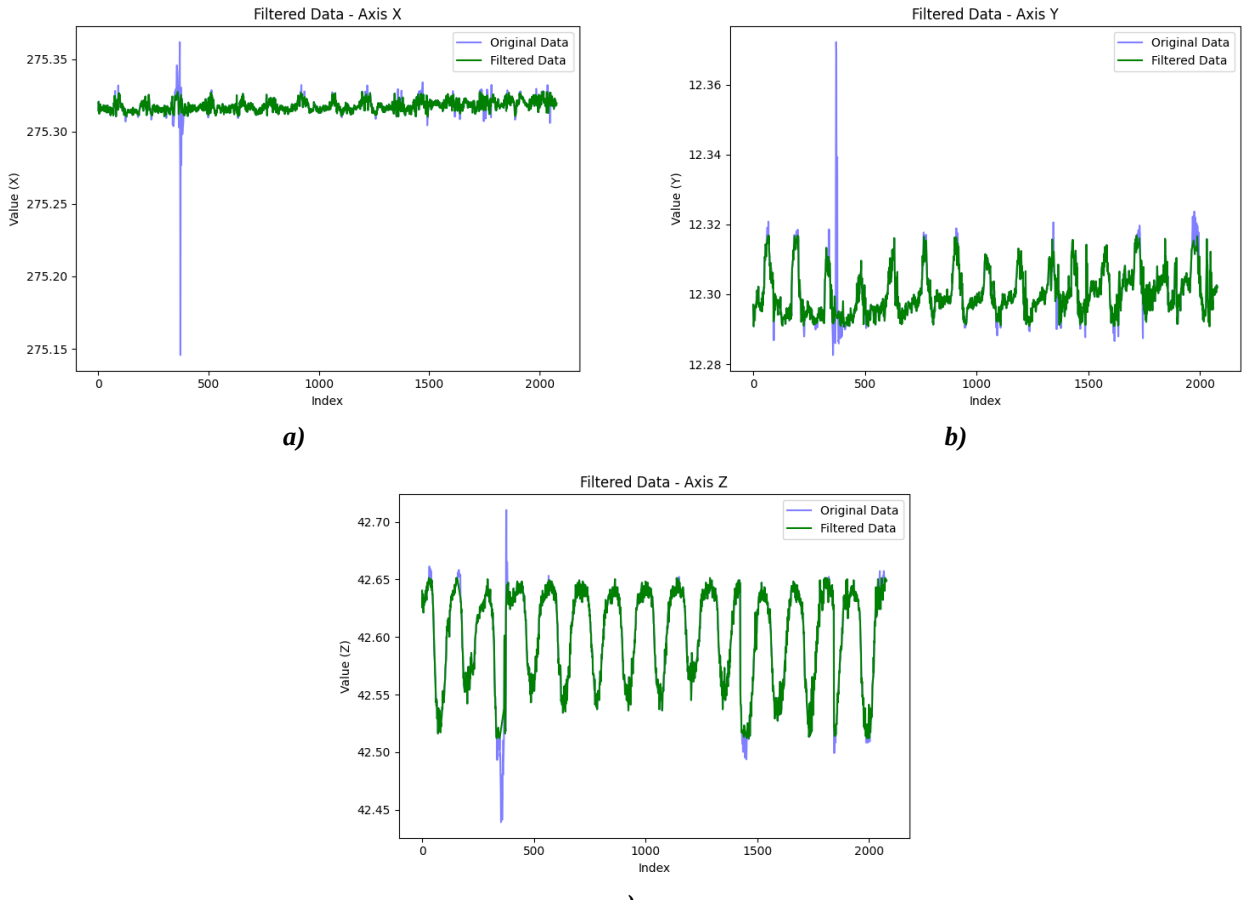
IF (Figure 4a) effectively captured significant vertical anomalies—notably around indices ≈ 400–450 and near the series end—concentrating on points with sudden amplitude changes or phase shifts. OC-SVM (Figure 4b) identified a somewhat larger set of anomalies, including lower-magnitude fluctuations at waveform crests and troughs, and produced a number of isolated detections in otherwise stable intervals. Thus, while both methods detect principal vertical disturbances, IF shows greater selectivity under the higher vertical noise typical of GNSS-RTK data.

3.2 Filtered GNSS-RTK Displacement Data after Outlier Removal

Outlier removal plays a critical role in SHM because GNSS-RTK measurements are highly sensitive to multipath effects, satellite geometry variations, and temporary signal obstructions. If such abnormal observations are not removed, they may be misinterpreted as structural deformation, leading to incorrect assessment of bridge safety or false damage alarms [6]. The purpose of the filtering process is therefore not to eliminate genuine structural movements, but to suppress measurement artefacts while preserving the physical deformation trend of the bridge. A successful outlier detection method should thus reduce signal dispersion without altering cyclic patterns, long-term displacement trends, or thermal responses.

Figure 5 compares the original (raw) and filtered time series for all three axes. The filtered signals (combined result of outlier exclusion) are smoother and more physically interpretable: abrupt spikes and irregular discontinuities present in the raw measurements are removed while overall deformation trends are preserved. On the X-axis (Figure 5a), the transient cluster near indices ≈ 400–450 is effectively suppressed in the filtered series. On the Y-axis (Figure 5b) the quasi-periodic lateral pattern remains intact but with isolated noise points eliminated. The Z-axis (Figure 5c) shows clearer diurnal cyclicity after filtering, revealing thermal-induced behavior that was partly obscured by high-amplitude noise in the raw data.

These visual improvements indicate that the proposed ML-based preprocessing enhances GNSS-RTK data interpretability without distorting genuine structural responses, making the filtered datasets more suitable for subsequent structural analyses (trend estimation, modal analysis, damage detection).



a) *b)*

c)

Fig. 5 - Comparison between the original and filtered GNSS-RTK displacement time series for the (a) X-axis, (b) Y-axis, and (c) Z-axis after outlier removal using the IF and OC-SVM.

3.3 Comparative Performance Evaluation

Table 2 summarizes the quantitative performance metrics computed for IF and OC-SVM on each displacement component. Evaluation metrics are: (1) number of detected outliers, (2) Silhouette Score (cluster separability), and (3) Reconstruction Error (fidelity of model representation). The count of overlapping outliers detected by both methods is also reported to assess consistency.

The IF consistently achieved higher Silhouette Scores across all axes (X: 0.678, Y: 0.571, Z: 0.453), indicating better separation between inlier and outlier clusters. IF also yielded substantially lower Reconstruction Errors (0.600–0.839) compared with OC-SVM (3.060–3.914), demonstrating superior fidelity in preserving legitimate displacement behavior after outlier identification. The number of detected outliers was comparable between methods (≈100 per axis), but OC-SVM tended to mark more low-amplitude deviations. The overlap counts (92, 100, 87) show a high degree of agreement—approximately 70–73% when related to the average detected outliers per axis—confirming that the majority of significant anomalies were jointly identified by both methods.

Table 2. Quantitative performance of IF and OC-SVM across GNSS-RTK components.

Algorithm	Metric	X	Y	Z
Isolation Forest	Number of outliers	101	104	104
	Silhouette Score	0.678	0.571	0.453
	Reconstruction Error	0.600	0.803	0.839
One-Class SVM	Number of outliers	105	105	98
	Silhouette Score	0.672	0.563	0.433
	Reconstruction Error	3.914	3.060	3.796
Number of overlapping outliers		92	100	87

Collectively, these quantitative results corroborate the visual assessments (Figures 2–5): Isolation Forest attains a more balanced detection pattern—minimizing false positives while preserving meaningful dynamics—whereas OC-SVM provides higher sensitivity to small fluctuations but at the cost of larger reconstruction errors.

3.4 Improvement in Data Quality after Filtering

Table 3. Statistical comparison of GNSS-RTK data quality before and after outlier filtering.

Algorithm	Axis	Std Before Filtering (m)	Std After Filtering (m)	Improvement (%)
Isolation Forest	X	0.005887	0.003283	44.24
	Y	0.007188	0.005829	18.90
	Z	0.043331	0.038095	12.08
One-Class SVM	X	0.005887	0.003298	43.98
	Y	0.007188	0.005895	17.98
	Z	0.043331	0.038776	10.51

Note: “Std Before Filtering” and “Std After Filtering” represent the standard deviations of displacement components before and after outlier removal, respectively. “Improvement (%)” denotes the relative reduction in data dispersion, reflecting enhanced signal stability after filtering.

Table 3 quantitatively compares the GNSS-RTK data quality before and after outlier removal using the Isolation Forest and One-Class SVM algorithms across all three displacement components (X, Y, and Z). Both methods significantly reduced

the standard deviation of the GNSS-RTK displacement signals, indicating effective noise suppression and improved data consistency.

The Isolation Forest achieved slightly higher improvements across all components, particularly along the longitudinal (X-axis) and transverse (Y-axis) directions, with reductions of 44.24% and 18.90%, respectively. The One-Class SVM yielded comparable performance, with improvements of 43.98% (X-axis) and 17.98% (Y-axis), though it exhibited slightly lower enhancement along the vertical (Z-axis).

Overall, both algorithms demonstrated robust capability in eliminating spurious GNSS-RTK observations without distorting the structural displacement trends. These improvements confirm the practical feasibility of applying unsupervised learning methods for real-time GNSS-based bridge monitoring, enhancing the stability and reliability of Structural Health Monitoring (SHM) systems.

3.5 Discussion on Practical Implications

The comparative analysis clearly demonstrates that both Isolation Forest and One-Class SVM algorithms substantially enhance the precision and stability of GNSS-RTK displacement data. These improvements validate the reliability of the proposed unsupervised learning framework and highlight its strong potential for real-world bridge monitoring applications.

By effectively removing outliers caused by multipath interference, atmospheric fluctuations, and temporary sensor malfunctions, the proposed framework preserves the true structural response of the bridge under normal operating conditions. Consequently, subsequent analyses - such as deformation trend identification, modal analysis, and vibration-based damage detection - can be performed using cleaner, more stable datasets.

From an engineering perspective, the observed reduction in displacement variability – up to 44% along the longitudinal (X) axis - demonstrates that the proposed framework significantly decreases false alarms and enhances the accuracy of maintenance decision-making. In practice, the system enables automatic identification of abnormal GNSS measurements, distinguishing them from genuine structural changes. This automation reduces manual post-processing and strengthens the reliability of long-term monitoring programs. Such functionality is particularly valuable for large-scale cable-stayed bridges like the Can Tho Bridge, where multiple sensors transmit continuous real-time data streams.

Moreover, since the proposed unsupervised learning methods do not require labeled training data, they are inherently suitable for real-time deployment within existing GNSS-RTK-based SHM systems. Once integrated into data acquisition platforms, the algorithms can continuously flag anomalies and dynamically update detection thresholds based on streaming data statistics – an essential capability for adaptive, intelligent SHM systems in the era of digital twins and smart infrastructure management.

The comparative findings also reveal practical trade-offs between the two algorithms. While Isolation Forest offers higher computational efficiency and robustness against non-Gaussian noise, One-Class SVM provides more precise boundary delineation when anomalies are densely clustered. Hence, a hybrid or ensemble approach combining both algorithms could yield an optimal balance between detection accuracy and real-time performance.

Overall, the comparative results confirm that the Isolation Forest algorithm is more suitable for GNSS-RTK outlier detection in bridge monitoring applications due to its greater robustness, lower computational complexity, and reduced dependence on hyperparameter tuning compared with the kernel-based One-Class SVM. These findings are consistent with previous studies in bridge monitoring. For example, Sun et al. (2023) applied IF to cumulative displacement data from a long-span cable-supported bridge in China and showed that the method effectively distinguishes structural anomalies from environmental effects [10]. Similarly, Bayane et al. (2024) used unsupervised learning algorithms, including IF and OC-SVM, for real-time damage detection in a monitored steel bascule railway bridge and demonstrated their robustness in identifying abrupt structural changes such as cracking events [17]. In addition, Mousavi and Gandomi (2022) validated the effectiveness of IF for unsupervised condition monitoring of the Z24 bridge using long-term frequency measurements, confirming its suitability for SHM under environmental variability [18].

By corroborating and extending these earlier results using real-world GNSS-RTK monitoring data from the Can Tho cable-stayed bridge, this study provides quantitative validation and a practical demonstration of how machine-learning–based anomaly detection can substantially improve the stability, interpretability, and long-term reliability of SHM systems.

4 Conclusions, suggestions, and limitations

4.1 Conclusions

This study developed and validated a machine-learning-based GNSS-RTK data processing framework for detecting and removing outliers in displacement monitoring of cable-stayed bridges. The experimental dataset consisted of 2,078 ten-minute averaged observations collected from the mid-span of the Can Tho cable-stayed bridge. The main conclusions are summarized as follows:

(1) Effectiveness of anomaly detection:

Both Isolation Forest (IF) and One-Class Support Vector Machine (OC-SVM) successfully identified abnormal measurements caused by temporary signal loss, multipath interference, and environmental disturbances. The two algorithms demonstrated high consistency, with more than 70% overlap in detected major anomalies.

(2) Comparative model performance:

The IF algorithm exhibited better cluster separability and reconstruction stability, achieving Silhouette Scores of X=0.678, Y=0.571, Z=0.453 and lower Reconstruction Errors (X=0.600, Y=0.803, Z=0.839). In contrast, OC-SVM produced comparable Silhouette Scores but much higher Reconstruction Errors (X=3.914, Y=3.060, Z=3.796), indicating a tendency to over-detect minor fluctuations.

(3) Improvement in data quality:

After outlier removal, both algorithms significantly enhanced the stability of the GNSS-RTK displacement series. The standard deviation was reduced by approximately 12%–44%, with the largest improvements observed along the longitudinal (X) and transverse (Y) axes. This confirms that the proposed framework effectively suppresses noise while preserving genuine structural responses.

(4) Practical significance:

Given its computational efficiency and superior robustness, the IF algorithm is suitable for integration as an automated preprocessing module in real-time SHM systems. A hybrid consensus strategy combining IF and OC-SVM can be employed to classify anomalies into high-confidence and review-required categories, thereby improving operational reliability.

4.2 Suggestions

Automated preprocessing integration: It is recommended that the IF algorithm be incorporated as a front-end module in GNSS-RTK data processing pipelines to automatically flag and filter outliers prior to structural analysis.

Consensus strategy: Adopt a hybrid decision rule - data points identified by both algorithms are removed automatically (high-confidence), while those detected by only one algorithm should be further verified using complementary sensor data (e.g., acceleration, temperature) or device logs.

Incorporating environmental information: Integrating environmental and traffic variables (temperature, wind speed, load) can improve the interpretation of anomalies and reduce false positives.

4.3 Limitations

Data scope: The analysis was limited to one GNSS-RTK monitoring point (mid-span) over approximately seven weeks, which may not capture seasonal or extreme operational events (e.g., typhoons, high traffic loads).

Parameter optimization: Model hyperparameters (e.g., contamination rate, ν, γ) were tuned empirically and may require recalibration for different bridge types or monitoring configurations.

Sensor diversity: This study relied solely on GNSS-RTK displacement data. The absence of multi-sensor integration (e.g., accelerometers, strain gauges, temperature sensors) restricts the physical interpretation of detected anomalies.

Data availability

The data sets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request. The authors declare that there are no restrictions on data access relevant to the findings of this research.

Declarations

Conflict of interest: The authors report there is no conflict of interest to declare.

REFERENCES

[1]- X.W. Ye, Y.H. Su, J.P. Han, Structural health monitoring of civil infrastructure using optical fiber sensing technology: A comprehensive review. Sci. World J., 1 (2014) 652329. https://doi.org/10.1155/2014/652329

[2]- A. Wieser, F.K. Brunner, Analysis of bridge deformations using continuous GPS measurements. In: Proceedings of 2nd International Conference on Engineering Surveying - INGEO, Bratislava, Slovakia, 2002, pp. 45-52.

[3]- K.G. Le, D.C. Tran, T.L.H. Ho, Self-supervised deep learning for GNSS time series imputation: a comparative study of neural network architectures. Transp. Commun. Sci. J., 77(1) (2026) 127-141. https://doi.org/10.47869/tcsj.77.1.10

[4]- F. Moschas, S. Stiros, Noise characteristics of high-frequency, short-duration GPS records from analysis of identical, collocated instruments. Measurement, 46(4) (2013) 1488-1506. https://doi.org/10.1016/j.measurement.2012.12.015

[5]- F. Moschas, S. Stiros, Dynamic deflections of a stiff footbridge using 100-Hz GNSS and accelerometer data. J. Surv. Eng., 141(4) (2015) 04015003. https://doi.org/10.1061/(ASCE)SU.1943-5428.0000146

[6]- J. Yu, X. Meng, X. Shao, B. Yan, L. Yang, Identification of dynamic displacements and modal frequencies of a medium-span suspension bridge using multimode GNSS processing. Eng. Struct., 81 (2014) 432-443. https://doi.org/10.1016/j.engstruct.2014.10.010

[7]- T. Kieu, B. Yang, C.S. Jensen, Outlier detection for multidimensional time series using deep neural networks. 19th IEEE International Conference on Mobile Data Management (MDM), (2018) 125-134. https://doi.org/10.1109/MDM.2018.00028

[8]- D. Yang, Y. Zhao, H. Ju, T.H. Yi, A. Li, Abnormal data detection for structural health monitoring: State-of-the-art review. Dev. Built Environ., 17 (2024) 100337. https://doi.org/10.1016/j.dibe.2024.100337

[9]- Z. Wang, Y.J. Cha, Unsupervised machine and deep learning methods for structural damage detection: A comparative study. Eng. Rep., 7(1) (2022) e12551. https://doi.org/10.1002/eng2.12551

[10]- Z. Sun, D.M. Siringoringo, S.Z. Chen, J. Lu, Cumulative displacement-based detection of damper malfunction in bridges using data-driven isolation forest algorithm. Eng. Fail. Anal., 143 (2023) 106849. https://doi.org/10.1016/j.engfailanal.2022.106849

[11]- H.T.L. Huong, T.D. Cong, L.V. Vu, L.K. Giang, Latent pattern recognition in GNSS-based SHM using t-SNE and adaptive time-series modeling. J. Civ. Struct. Health Monit., 15 (2025) 3743–3766. https://doi.org/10.1007/s13349-025-01014-9

[12]- H.T.L. Huong, T.D. Cong, L.V. Hien, L.K. Giang, A study on the application of unsupervised clustering algorithms in GNSS-RTK data analysis for cable-stayed bridges monitoring. Transp. Commun. Sci. J., 76(8) (2025) 1138-1150. https://doi.org/10.47869/tcsj.76.8.8

[13]- N. Shen, B. Wang, G. Gao, L. Chen, 3-D Displacement Detection Based on Enhanced Clustering From GNSS Positioning in a Kinematic Mode for Deformation Monitoring. IEEE Trans. Instrum. Meas., 72 (2022) 1-10. https://doi.org/10.1109/TIM.2022.3223072

[14]- J. Long, O. Buyukozturk, Automated structural damage detection using one-class machine learning. In: Dynamics of Civil Structures, Volume 4: Proceedings of the 32nd IMAC, A Conference and Exposition on Structural Dynamics, 2014, pp. 117-128. https://doi.org/10.1007/978-3-319-04546-7_14

[15]- O. Çetindemir, E. Tepe, A.C. Zülfikar, A. Yeşilyurt, N.M. Apaydın, Long-term structural health monitoring of long-span suspension bridges and anomaly detection using statistical indicators. In: Proceedings of the 11th International Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII-11), Turin, Italy, 2022, pp. 241-244.

[16]- N.L. Khoa, B. Zhang, Y. Wang, F. Chen, S. Mustapha, Robust dimensionality reduction and damage detection

approaches in structural health monitoring. Struct. Health Monit., 13(4) (2014) 406-417. https://doi.org/10.1177/1475921714532989

[17]- I. Bayane, J. Leander, R. Karoumi, An unsupervised machine learning approach for real-time damage detection in bridges. Eng. Struct., 308(1) (2024) 117971. https://doi.org/10.1016/j.engstruct.2024.117971

[18]- M. Mousavi, A.H. Gandomi, Unsupervised condition monitoring of structures using VMD and isolation forest. Struct. Health Monit., (2021). https://doi.org/10.1177/14759217211013532

[19]- V.H. Le, D.H. Vo, A case study of processing abnormal GNSS monitoring data of a cable-stayed bridge in Vietnam. IOP Conference Series: Materials Science and Engineering, 1289(1) (2023) 012035. https://doi.org/10.1088/1757-899X/1289/1/012035

[20]- T.T. Duong, N.Q. Long, B.V. Duc, Hybrid-Precision GNSS Positioning Strategies for Landslide Monitoring. Int. J. Environ. Sci., 11(13s) (2025) 11–22. https://doi.org/10.64252/pfbba722

[21]- T.D. Tran, T.D. Dao, T.S. Vu, D.N. Luong, C.D. Vu, S.B. Bui, H.T. Ha, Outlier detection in GNSS position time series. Sci. Technol. Dev. J., 19(2) (2016) 43-50. https://doi.org/10.32508/stdj.v19i2.665

[22]- H.D. Nguyen, T.D. Tran, Detecting outliers in GNSS position time series using machine learning techniques. J. Min. Earth Sci., 64(4) (2023) 22–30. https://doi.org/10.46326/JMES.2023.64(4).03

[23]- F.T. Liu, K.M. Ting, Z.H. Zhou, Isolation forest. 2008 Eighth IEEE International Conference on Data Mining, (2008) 413-422. https://doi.org/10.1109/ICDM.2008.17

[24]- Y. Regaya, F. Fadli, A. Amira, Point-Denoise: Unsupervised outlier detection for 3D point clouds enhancement. Multimed. Tools Appl., 80 (2021) 28161-28177. https://doi.org/10.1007/s11042-021-10924-x

[25]- B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, R.C. Williamson, Estimating the support of a high-dimensional distribution. Neural Comput., 13(7) (2001) 1443-1471. https://doi.org/10.1162/089976601750264965

[26]- C.C. Chang, C.J. Lin, LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol., 2(3) (2011) 1-27. https://doi.org/10.1145/1961189.1961199

[27]- L.M. Manevitz, M. Yousef, One-class SVMs for document classification. J. Mach. Learn. Res., 2 (2001) 139-154.

[28]- L.K. Giang, H.H.T. Lan, D.V. Manh, T.Q. Hoc, Applying a two-step cluster algorithm in traffic accident data analysis. Transp. Commun. Sci. J., 75(4) (2024) 1673-1687. https://doi.org/10.47869/tcsj.75.4.16

[29]- K.R. Shahapure, C. Nicholas, Cluster quality analysis using silhouette score. 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), (2020) 747-748. https://doi.org/10.1109/dsaa49011.2020.00096

[30]- L.H. Viet, T.T. Thi, B.H. Xuan, Swarm intelligence-based technique to enhance performance of ANN in structural damage detection. Transp. Commun. Sci. J., 73(1) (2022) 1-15. https://doi.org/10.47869/tcsj.73.1.1

[31]- AASHTO, AASHTO LRFD Bridge Design Specifications. 9th Edition, American Association of State Highway and Transportation Officials (AASHTO), 2020.