Tunnelling and Underground Space Technology 169 (2026) 107264

Contents lists available at ScienceDirect

Tunnelling and Underground Space Technology incorporating Trenchless Technology Research

journal homepage: www.elsevier.com/locate/tust

Trenchless technology research

Recurrent normalizing flow-based monitoring framework for cutter seal temperature sensors in earth pressure balance shield tunneling: A sensor validation approach

Jorge Loy-Benitez, Je-Kyum Lee, Myung Kyu Song, Fabian Cabrera Guerra, Sean Seungwon Lee *

Department of Earth Resources and Environmental Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea

ARTICLE INFO

Keywords:
Earth pressure balance
LSTM autoencoder
Recurrent autoregressive flow
Shield tunneling
Sensor validation

ABSTRACT

Reliable monitoring of cutter seal temperature is essential for the safe and efficient operation of earth pressure balance (EPB) machines. Faulty temperature sensors can mislead operators, potentially leading to seal degradation, slurry or water ingress, and costly downtime. This study introduces a data-driven monitoring framework that automatically detects and reconstructs faulty temperature sensor readings to ensure robust information during tunneling. The proposed method combines a long short-term memory-based autoencoder with a recurrent autoregressive flow, namely RAF-LSTM, to learn normal temperature patterns over time and model the probability of those patterns to distinguish normal from abnormal behavior. The proposed method was assessed using operational data from an EPB project, where synthetic faults representing realistic sensor failures have been introduced. Results show a fault detection accuracy of 96 % with 100 % recall and reconstruction errors below 0.6 °C. These outcomes demonstrate that the proposed framework can reliably identify and correct faulty sensor signals, reducing false alarms and supporting proactive maintenance strategies for EPB machine operations.

1. Introduction

Tunnel-boring machines (TBMs) and shield excavation with Earth Pressure Balance (EPB) machines have recently emerged as preferred tunnel engineering methods for urban environments due to their efficiency, cost-effectiveness, safety, and minimal disturbance (Yin et al., 2024). However, the operational performance of TBMs depends heavily on continuous monitoring of both machine components and ground responses, affecting the excavation efficiency, budget, and duration of a tunneling project (Pan et al., 2022; Wang et al., 2018). Faulty or unreliable sensor measurements can mislead operators, potentially causing undetected failures, unexpected downtime, or costly repairs. As tunneling projects become increasingly data-rich, there is a growing interest in adopting advanced monitoring methods to ensure robust decision-making (Akhlaghi et al., 2025).

With the emergence of Industry 4.0 and Construction 4.0, various artificial intelligence (AI) methods, IoT-enabled sensing, and digital twin technologies into construction and tunneling (Gomaa et al., 2023; Tariq et al., 2024). These advances created opportunities for predictive maintenance, automated quality control, and process optimization in underground works. Recent studies have demonstrated the value of smart monitoring across a broad range of tunneling applications. For instance, disc cutter wear and lifetime have been predicted using neural approaches (Elbaz et al., 2021), while gripper cylinder faults have been identified through vibration analysis and signal processing methods (Li et al., 2017). On the other hand, when monitoring ground responses, machine learning (ML) approaches have been widely applied to predict surface settlement and detect sinkhole formations (Liu et al., 2024; Loy-Benitez et al., 2024). These investigations illustrate the potential of data-driven methods to support machine reliability and ground safety. Sensors are a valuable tool for monitoring abnormal events during the operation of TBMs and ensuring the integrity of the constituent elements by sounding an alarm when maintenance or replacement is necessary (Wang et al., 2019). However, during excavation, strong and continuous vibrations and the harsh tunneling conditions (e.g., wet and highly pressurized) can damage these sensors, leading to biased information or sensor drift. To address this issue, it is useful to integrate sensor information with emerging data-driven techniques to identify faulty sensors

* Corresponding author.
E-mail address: seanlee@hanyang.ac.kr (S.S. Lee).

https://doi.org/10.1016/j.tust.2025.107264
Received 2 July 2025; Received in revised form 23 September 2025; Accepted 7 November 2025
Available online 18 November 2025

as part of a fault detection, fault diagnosis, and reconstruction framework, a strategy that has proven successful for wastewater treatment, mechanical, and HVAC systems (Ba-Alawi et al., 2021; Wang et al., 2019).

In this context, sensor validation is a data-based process monitoring technique that can be used to identify faulty sensors. This method examines a range of sensor signals to detect, identify, and holistically reconstruct faulty sensors (Wang and Chen, 2004). Principal component analysis (PCA) is used as a baseline for sensor validation, learning the latent space from a set of data points obtained under normal operating conditions via linear transformations (Liu et al., 2013). However, this method struggles to handle multiple faults, complex failures, and non-linear relationships (Brooks and Bauer, 2018). For this reason, various approaches based on non-linear kernels and dynamic windows have been developed, including kernel PCA (KPCA) and dynamic PCA (DPCA) (Han et al., 2021; Navi et al., 2018a). Neural approaches have also been used as a substitute for PCA, replacing linear transformations with non-linear activations in autoencoders (AEs). These neural networks have been improved by extending their memory cells for sequential data processing, increasing their depth to capture data nonlinearities, and incorporating adversarial training mechanisms for the detection of minor anomalies (Kang et al., 2024; Loy-Benitez et al., 2020; Yan et al., 2016). Despite their advantages, these methods still face challenges related to overfitting, limited temporal modeling capacity, and the absence of robust probabilistic scoring mechanisms.

Particularly, the integrity of the bearings and seals is critical for the reliable and safe operation of TBMs. Bearings are found within the cutterhead and the main drive system. Taper roller bearings allow the cutting discs to rotate and are subject to high transient loads during operation, while slewing bearings link the pinion drives and the cutterhead, thus supporting very high loads. The fault-free operation of the slewing bearings relies on proper lubrication and protection against ingress from excavation materials and water. Therefore, high-performance seals are required to protect the bearings during TBM excavation (SKF, 2019). Monitoring these seals helps to maintain the function of the bearings, ensuring the reliable operation of the TBM. This is particularly important because these machines do not have a reverse gear; thus, excavation must be completed without failure.

Temperature is a key indicator of the status of sealed bearing systems. It is generally measured with thermocouples or resistance temperature detectors embedded near or within the seal housing, and the data is logged and displayed in the control room (Fu et al., 2021). The temperature of the seals in the cutters can be used for the early detection of failure because, if the seal wears or breaks, contaminants can enter, increasing friction and subsequently the temperature in the bearings and cutterhead. This leads to both cutter abrasion and thermal stress, which affects the viscosity of the soil, forming mud cakes (Kalayci Sahinoglu and Ozer, 2020). Trends in the temperature can also be used to determine when the cutters need to be replaced, while overheating can also indicate the breakdown of lubrication. However, the harsh tunneling environment can damage temperature sensors; thus, providing incorrect information to the control room and leading to unnecessary actions or preventing timely responses.

Generative models in monitoring frameworks are based on latent representations that accurately reconstruct data collected under normal conditions (Zhao et al., 2024). However, AEs remain susceptible to overfitting, reducing their reconstruction accuracy. In addition, as the AE structure becomes more complex, the models memorize patterns rather than learning from the latent representation (Pang et al., 2021). To overcome these issues, normalizing flows (NFs) have been employed because of their advantages over AEs and other generative models, including the generation of tractable distributions that allow for precise anomaly scoring (Kobyzev et al., 2021). The invertible NF architecture also enhances the interpretability of deviations from normal behavior, while flexible density estimation captures complex patterns for the detection of subtle anomalies.

Due to the temporal dependencies present in process data, recurrent NFs (RNFs) have been constructed to combine the likelihood estimation capabilities of a regular NF with temporal correlations, with anomaly detection conducted without relying on reconstruction errors (Mern et al., 2020). On the other hand, the reconstruction ability of memory cell-based architectures, such as long short-term memory (LSTM) or gated recurrent unit (GRU) models, has been demonstrated to be superior to baseline methods due to their sequential data processing (Qin et al., 2023). Therefore, this study introduces a sensor validation approach for the detection and reconstruction of faulty cutter seal temperature sensors using a hybrid framework that integrates an LSTM-based AE for temporal reconstruction with an RNF for latent space probabilistic modeling. The proposed model thus exploits both the reconstruction ability of the recurrent network and the density estimation capability of the RNF to improve the fault detection performance of the model.

This paper is organized as follows. Section 2 presents a summary of recurrent AEs, NFs, and RNFs. Section 3 then describes the tunneling project, data collection, proposed experimental methodology, and corresponding performance metrics. Following this, Section 4 presents the model training process and modeling results with a discussion related to the implications on the tunneling operations. Finally, a conclusion is provided in Section 5.

2. Preliminaries

2.1. Recurrent autoencoders

AEs are unsupervised machine learning models consisting of two components: 1) an encoder that learns a compressed representation of a given input, and 2) a decoder that reconstructs the input from the obtained representation (Yan et al., 2016). The generalized AE structure is presented in Fig. 1(a), where input $X \in \mathbb{R}^d$ is mapped to latent representation $h \in \mathbb{R}^d$ using Eq. (1):

$$h = f(x) = \sigma_f(Wx + b) \tag{1}$$

where W represents the $d' \times d$ weight matrix, $b \in \mathbb{R}^{d'}$ is the bias vector, and $\sigma_f(\cdot)$ is the non-linear activation function.

The decoder reconstructs the input by mapping the latent representation space, forcing the AE to select the representation that reconstructs the input as accurately as possible. This process is summarized as Eq. (2):

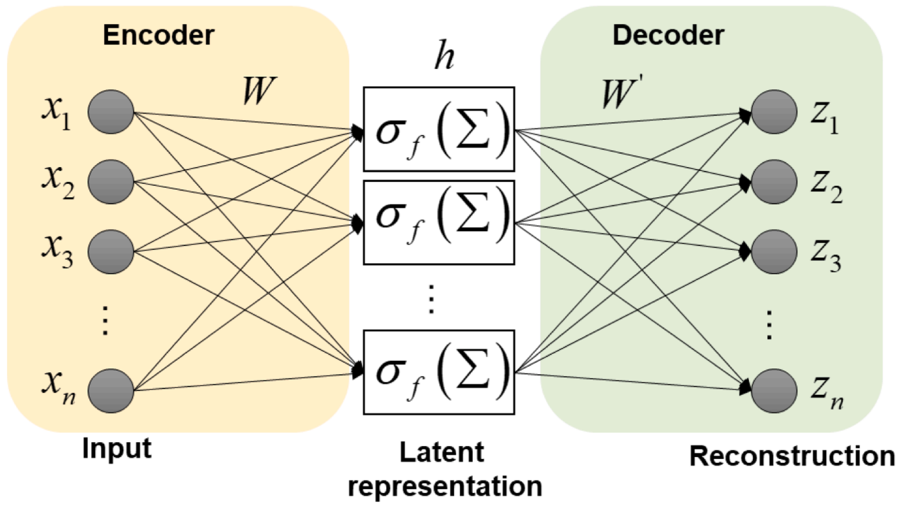
$$z = g(h) = \sigma_g(W'h + b') \tag{2}$$

where $\sigma_g(\cdot)$ is the activation function for the decoder and the parameters are $d \times d'$ weight matrix $W' = W^T$ and bias vector $b' \in \mathbb{R}^d$. The AE model is trained using backpropagation to minimize the reconstruction error between the input and its reconstruction.

AEs are generally used to address non-linear relationships in a dataset due to their superiority over state-of-the-art PCA for process monitoring, but these models cannot reliably handle temporal correlations in the data. Therefore, recurrent layers have been introduced because they can control data propagation with gate units and memory cells that retain or discard information (Cortez et al., 2018). The present study employs an LSTM-based AE, with the encoding/decoding process computed via recurrent operations in recurrent neural network (RNN) memory cells (Fig. 1b). These cells consist of three main gates: the input gate (i_t), the output gate $\left(\widetilde{o}_t\right)$, and the forget gate $\left(\widehat{f}_t\right)$, where $\breve{c}_t$ is the candidate value (Hochreiter and Schmidhuber, 1997).

$$i_t = \delta\left(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i\right) \tag{3}$$

$$\widetilde{o}_t = \delta\left(W_{\widetilde{o},x}x_t + W_{\widetilde{o},h}h_{t-1} + b_{\widetilde{o}}\right) \tag{4}$$



(a)

Fig. 1. Schematic of a (a) general AE structure and (b) memory cell-based AE structure containing LSTM cells.

$$\hat{f}_t = \delta\left(W_{\hat{f},x}x_t + W_{\hat{f},h}h_{t-1} + b_{\hat{f}}\right) \tag{5}$$

$$\breve{c}_t = \tanh\left(W_{\breve{c},x}x_t + W_{\breve{c},h}h_{t-1} + b_{\breve{c}}\right) \tag{6}$$

where $W_{\cdot,x}$ and $W_{\cdot,h}$ are weight matrices, $b_{(\cdot)}$ represents the bias vector, x_t is the current input, h_{t-1} is the output of the LSTM cell at time *t-1*, and $\delta(\cdot)$ is the activation function.

In the LSTM-based AE, the following transitions occur:

(i) $\hat{f}_t$ dictates how much of the previous memory is removed from the cell state (c_t).

(ii) i_t provides the new input to c_t, as shown in Eq. (7):

$$c_t = \hat{f}_t \circ c_{t-1} + i_t \circ \breve{c}_t \tag{7}$$

where $\circ$ is the element-wise Hadamard product.

(iii) The latent representation obtained from the encoding process in the LSTM cell is as follows:

$$h_t = \tilde{o}_t \circ \tanh(c_t) \tag{8}$$

(iv) Finally, h_t is projected to reconstruction z_t, using Eq. (9):

$$z_t = W_z h_t \tag{9}$$

where W_z is the projection matrix that reduces the dimensionality of h_t.

2.2. Recurrent normalizing flows

NFs estimate complex probability distributions by transforming base distributions such as simple Gaussians (Kobyzev et al., 2021). This process is conducted through invertible transforms with tractably assessable Jacobian determinants, which are sufficiently flexible to model multi-modal and complex densities. However, these models generally utilize highly parameterized function approximators such as neural networks, meaning that modeling conditional distributions limits the learned transforms (Dinh et al., 2017). Therefore, Mern et al. introduced RNFs as a flexible mechanism that uses bijective recurrent cells based on GRUs for recurrent autoregressive flows (RAFs), conditioning the transform to previous timesteps (Mern et al., 2020).

An NF consists of function $Z = f(X)$, which transforms a random variable to another and meets the following criteria: the inverse function $X = f^{-1}(Z)$ exists, the Jacobian determinant $\left|\frac{\partial f(x)}{\partial x}\right|$ is assessable, and bijection is obeyed in a way that, for all $x \in X$, *f(X)* maps to only one z and vice versa.

The probability distribution for X transformed from variable Z via the NF is expressed as follows:

$$p_X(x) = p_z(f(x))\left|\det\left(\frac{\partial f(x)}{\partial x}\right)\right| \tag{10}$$

The learning process for the NF consists of adjusting f to maximize the likelihood of the observed data.

RNFs consider vector $X = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and, instead of independent modeling, they model each component conditionally utilizing hidden state h_t, which is recurrently updated in a RAF cell using conditioning vector $z_t = f(X_t, \theta(X_{t-1}))$, as illustrated in Fig. 2(a). These cells contain the conditioning cell and the transformer function. The conditioning cell is based on a GRU, using gate functions to control the progress of the hidden state over time. At each time step, h_t, is a function of the observed input X_t and h_{t-1}. The GRU cell has two main gates: a reset gate (r_t) that is applied to previous states and an update gate (ν_t). These are analogs to the input and output gates of LSTM. Given *Xt* as the observed input, the GRU transition is illustrated in Fig. 2(b) and described in Eqs. (11)–(14) (Cho et al., 2014):

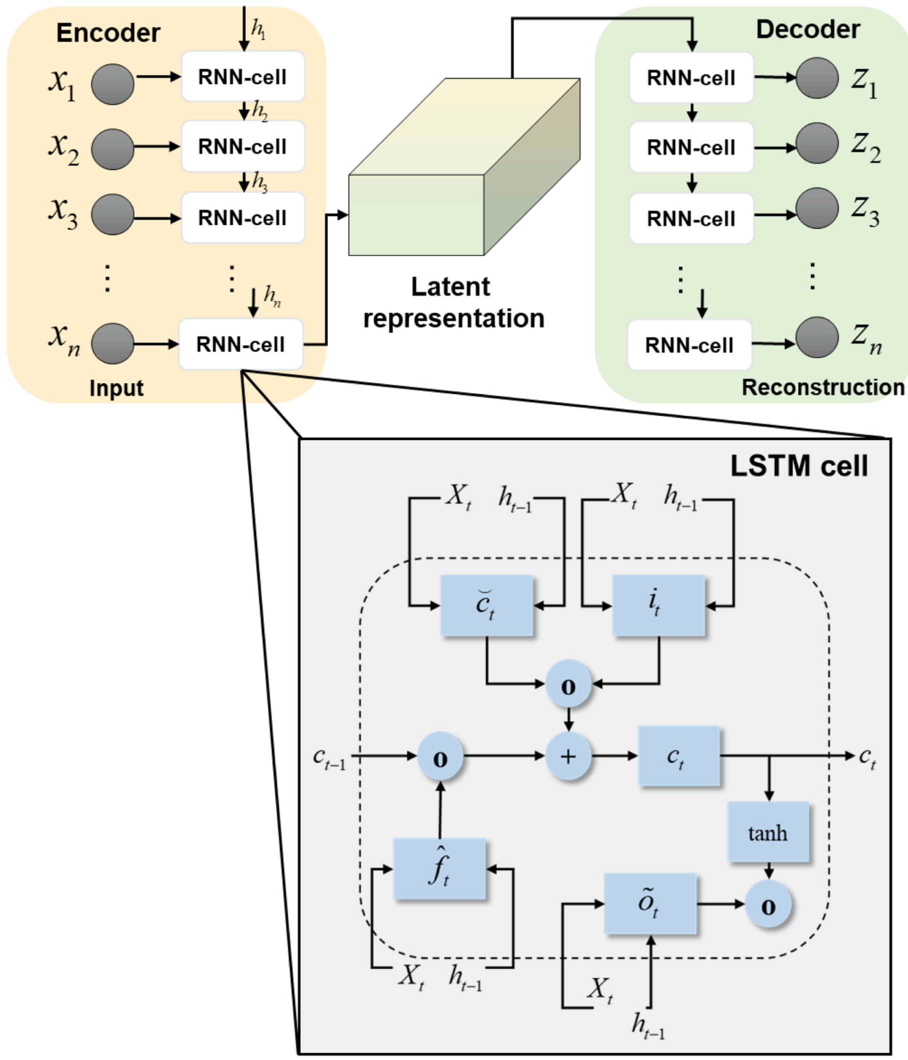
$$\nu_t = \phi(W^{\nu}X_t + U^{\nu}h_{t-1}) \tag{11}$$

$$r_t = \phi(W^{r}X_t + U^{r}h_{t-1}) \tag{12}$$

$$h'_t = \tanh(WX_t + r_t \circ Uh_{t-1}) \tag{13}$$

$$h_t = (1 - \nu_t)^{\circ} h'_t + \nu_t{}^{\circ} h_{t-1} \tag{14}$$

where W^{ν}, U^{ν}, W^{r}, U^{r}, W, U are weight matrices, $\phi(\cdot)$ is the activation function, and h'_t and h_t are the memory contained in the cell and the final hidden state memory, respectively. ν_t controls how much data is propagated from the previous storage, and r_t integrates the new data



(b)

Fig. 1. (*continued*).

points with the stored information.

The transformer function, which can be set as any bijective flow, considers obtained hidden state vector h_t and splits it into a weight vector (h_t^w) and a bias vector (h_t^b). h_t^w is then transformed into a lower-diagonal matrix (W_t^h), that, with the bias vector, is applied to an affine transform and an invertible non-linear function $f(\cdot)$ to obtain the transform shown in Eq. (15):

$$Z_t = f\left(W_t^h X_t + h_t^b\right) \tag{15}$$

3. Materials and methods

3.1. Data description and preliminary analysis

The TBM considered in this study is an EPB used in the metropolitan area of Seoul (Table 1). The tunnel is 1,600 m long with two stations, seven ventilation shafts, and an alignment length of 760 m and 762 m for the northern and southern tunnels, respectively. This project consists of 1,011 segment rings with a total excavation length of 1,522 m. The tunnel is approximately 6 to 20 m deep, and the ground is a mix of weathered and soft rock with an alluvial zone at the arrival. Groundwater level is encountered 7–8 m below the surface and 3–4 m above the tunnel crown. The geological layout is illustrated in Fig. 3, while the geological characteristics are summarized in Table 2.

The TBM-EPB operational data consists of 42 variables with 75–80 measurements per ring at time intervals of 2–3 min. Table A1 in Appendix A lists the parameters measured during TBM-EPB operation and their descriptive statistics. Because the present study employs models that consider temporal dependencies in the data, a preliminary analysis using the Hurst exponent (H) is conducted for the TBM-EPB data. H ranges between 0 and 1; when $H = 0.5$, the time series follows Brownian motion or is independent, when $0 \leqslant H < 0.5$, the time series is considered anti-persistent, and when $0.5 < H \leqslant 1$, the time series is classified as persistent. H is derived from rescaled range analysis (R/S) based on Eq. (16) (Lotfalinezhad and Maleki, 2020):

$$\log(R/S)_n = \log c + H \log n \tag{16}$$

where R is the range, S is the standard deviation, n is the length of the subseries, and c is a positive constant.

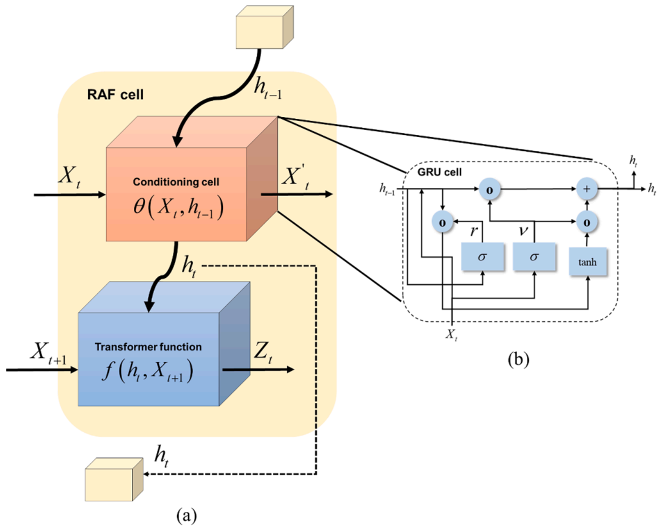


Fig. 2. Structure of the (a) recurrent normalizing flow cell and (b) conditioning GRU memory cell.

Table 1
Characteristics of the TBM-EPB utilized in the present study.

Item	Description
Type/supplier	Earth pressure balance/Hitachi Zosen (Japan)
Grouting	Upper section 4EA, Lower section 2EA, Chamber 4EA
Muck handling	Muck car + vertical conveyor belt, belt scale
ID/OD	7.47 m/7.69 m
Segment	RC-segment, L 1,500 mm + t 300 mm, 7 pieces
RPM	0.6–3.53 RPM, electric motor type
Torque	11,300–43,300 kN-m (α = 24.8)
Cutter head	Dome type, 17-inch cutter, scraper
Thrust force	1,120 kN/m^2/52,000 kN

3.2. Sensor validation for faulty cutter seal temperature sensors

The present study aimed to test a monitoring framework for the detection, identification, and reconstruction of faulty measurements during the tunneling process. The multiple seal temperature sensors in the TBM-EPB are monitored using the RAF approach. A total of 7,936 observations are collected and divided into training and test sets (70 % and 30 %, respectively). The training set is used to establish the normal process conditions to be learned by the monitoring model, while the test set is employed to validate the performance of the model. Because the measurements were obtained under normal conditions, artificial faults are introduced to the test set to evaluate the detection and reconstruction performance. Based on the sensor hardware, four common fault types are introduced to the test set: complete failure, bias failure, drift, and precision degradation (Liu et al., 2012).

This study introduces synthetic faults, which are a common practice in fault detection and sensor validation research (Ba-Alawi et al., 2021; Navi et al., 2018b). The use of artificial faults enables systematic testing of detection and reconstruction capabilities when real-world faulty data are scarce or not accessible due to project confidentiality or risks and costs associated with TBM operation under abnormal conditions. In addition, the selected failure types represent typical scenarios reported in the literature and allow controlled evaluation of the proposed framework. The type of failure, its magnitude, and the sensors in which the fault occurs are summarized in Table 3 and illustrated in Fig. 4(a), while Fig. 4 (b)–(e) present violin plots showing the distribution for the normal (left), and faulty (right) measurements.

3.2.1. RAF-LSTM architecture

The sensor validation framework in this study uses an RAF-LSTM architecture consisting of two phases (Fig. 5). Phase 1 trains the recurrent AE with the aim of accurately reconstructing the operational TBM-EPB in a latent space (H), while Phase 2 estimates the distribution density of this latent space.

Phase 1 – Recurrent AE: The LSTM-based recurrent AE encodes the multivariate dataset into a latent space, preserving contextual information that can then be decoded to produce a reconstruction that accurately models the input (Qin et al., 2023). In order to monitor the seal temperature sensors, this model learns both the behavior of the data under normal conditions and how to reconstruct it from the latent space. The recurrent AE contains an input layer with 42 nodes, corresponding to the number of features in the dataset, and a hidden layer or hidden space with 13 nodes, corresponding to the number of principal components obtained from the PCA that explains 85 % of the variance. This

Table 2
Geological information for the tunneling project.

Item	Rock characteristics
Residual soil	Sandy gravel, N: 3–50, max size: φ500 mm
Weathered rock	Silty core, Cohesion: 31 kPa, φ = 32°
Soft rock	Gneiss, RMR: 30–50
Water level	Approx. GL-7 m
Unit weight	20–21 kN/m^2
UCS	20–110 MPa
Permeability	2.5 × 10^{-3}–2.5 × 10^{-5} cm/s

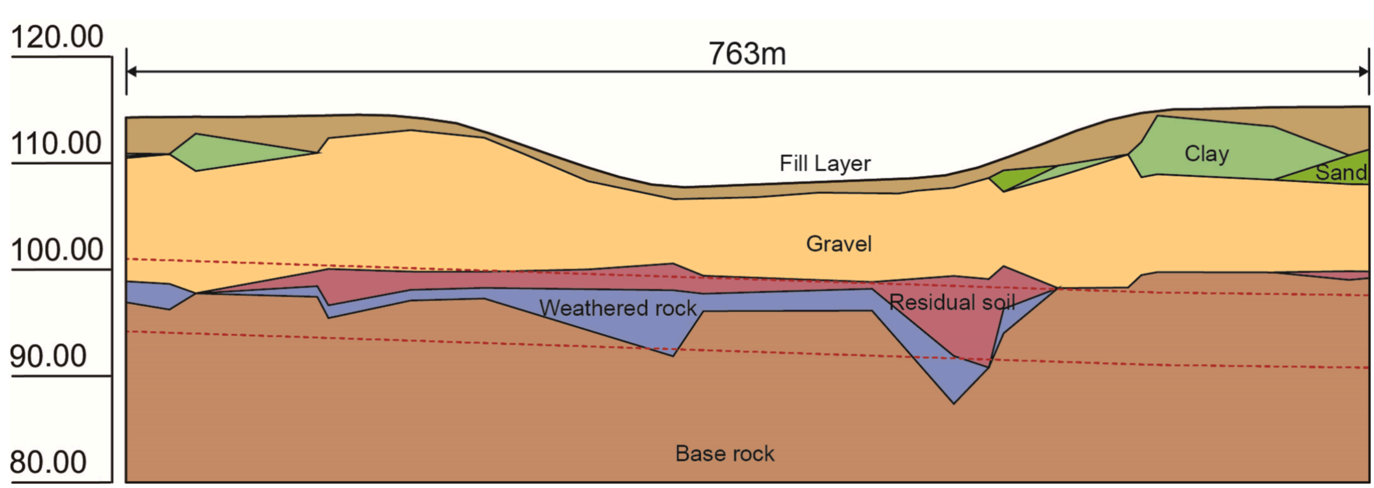
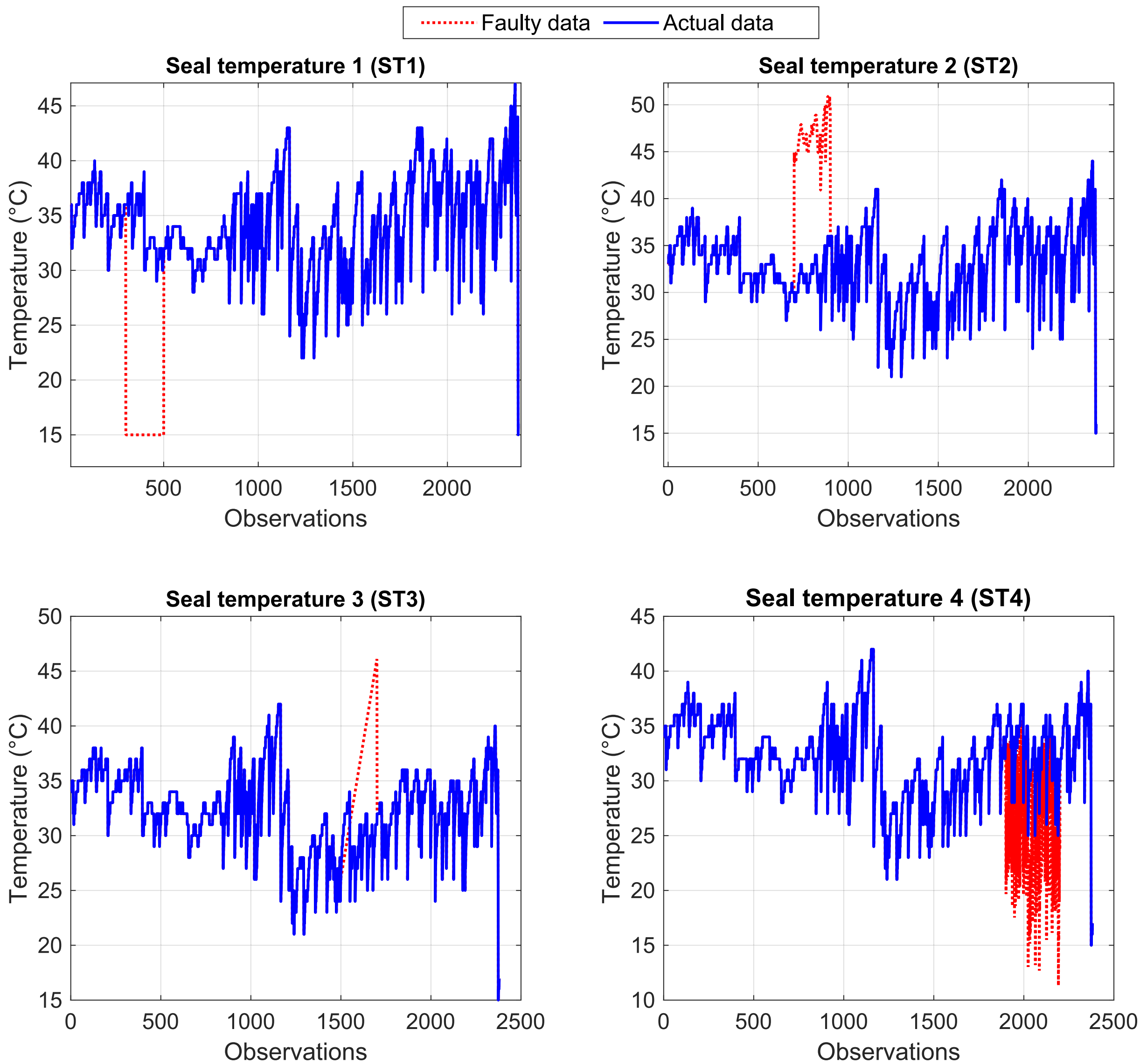


Fig. 3. Geological layout of the focal tunneling project.

Table 3
Faulty measurements introduced to the seal temperature sensors.

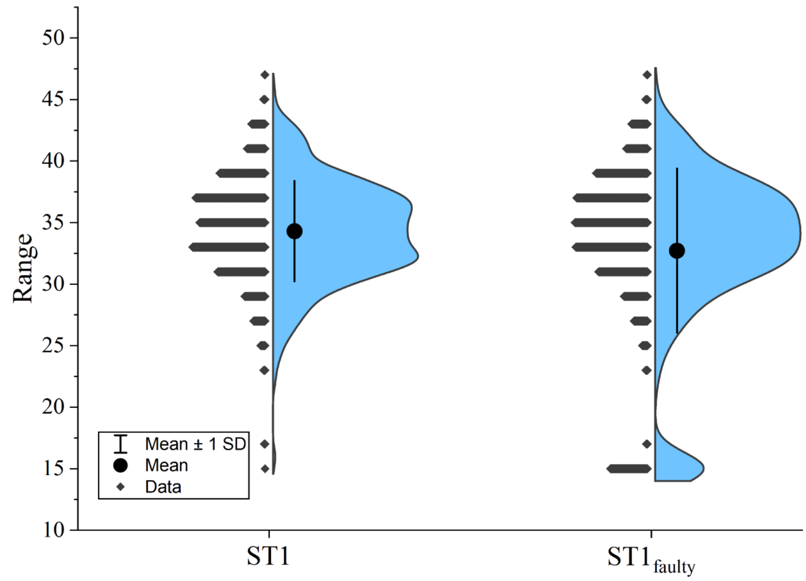
Fault type	Normal	Complete failure	Bias	Drifting	Precision degradation
Faulty sensor		ST1	ST2	ST3	ST4
Fault magnitude		15 ℃	47 ℃	0.1 ℃/sample	rand(0.5μ)
Fault occurrence period	1–299, 501–699, 901–1499, 1701–1899, 2201–2381	300–500	700–900	1500–1700	1900–2200



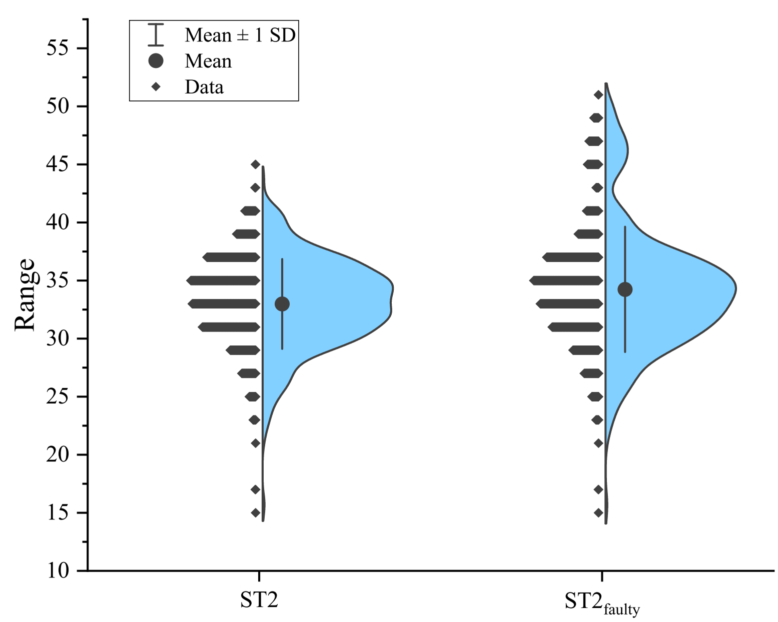
(a)

Fig. 4. Faulty measurements introduced to the seal temperature test set, including (a) four common fault types for each temperature sensor and the distribution before and after the introduction of the faults for sensors (b) ST1, (c) ST2, (d) ST3, and (e) ST4.

process is described in more detail in Section 1 of the Supplementary Information (SI). The output layer subsequently stores the reconstructed data in 42 nodes. This model is trained by minimizing the reconstruction error using the mean squared error (MSE) as the loss function (Eq. (17):

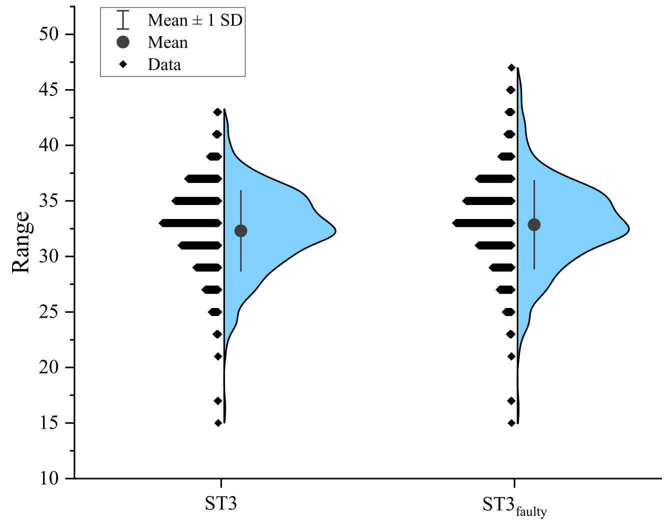


(b)

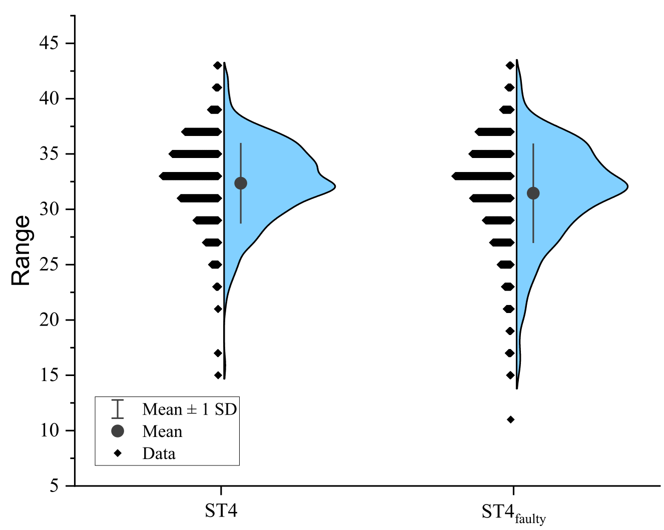


(c)

Fig. 4. (*continued*).



(d)



(e)

Fig. 4. (*continued*).

$$\ell = \underset{\phi,\theta}{\operatorname{argmin}} \sum_{i=1}^{N} \left(X_i - \underbrace{g_\theta\left(f_\phi(X_i)\right)}_{\widetilde{X}_i} \right)^2 \tag{17}$$

where ϕ and θ are the parameters to be adjusted after feedback optimization for the encoder $f(\cdot)$ and decoder $g(\cdot)$, respectively.

The neural structure is generated using the Keras deep learning package with the following hyperparameters:

- Optimizer: Adam; learning rate $= 0.001$
- Batch size: 64
- Activation function from X to H: tanh
- Activation function from H to $\widetilde{X}$: linear
- Epochs: 10,000; early-stopping: patience – 300 epochs

Phase 2 – RNF: While the recurrent AE encodes a p-dimensional latent space ($p = 13$), the RNF is trained to learn this complex distribution. The RAF model generates invertible transformations $f_1, f_2, \ldots, f_K$, which are sequentially applied to standard Gaussian distribution Z. These transformations convert the Z distribution into the complex distribution of the latent space (Eq. (18):

$$H = f_K{}^{\circ} f_{K-1}{}^{\circ} \ldots {}^{\circ} f_1(z) \tag{18}$$

where $z \sim N(0, I)$ and $f(\cdot)$ are the recurrent transformations for conditioning and the transformer cells, respectively (see Section 2.2). The *nflows* package developed with the PyTorch framework is used for the computation (Durkan et al., 2020). The RAF is customized with GRU cell parameterization for affine transformations, which are wrapped in a flow with base distribution Z. The training hyperparameters are as follows:

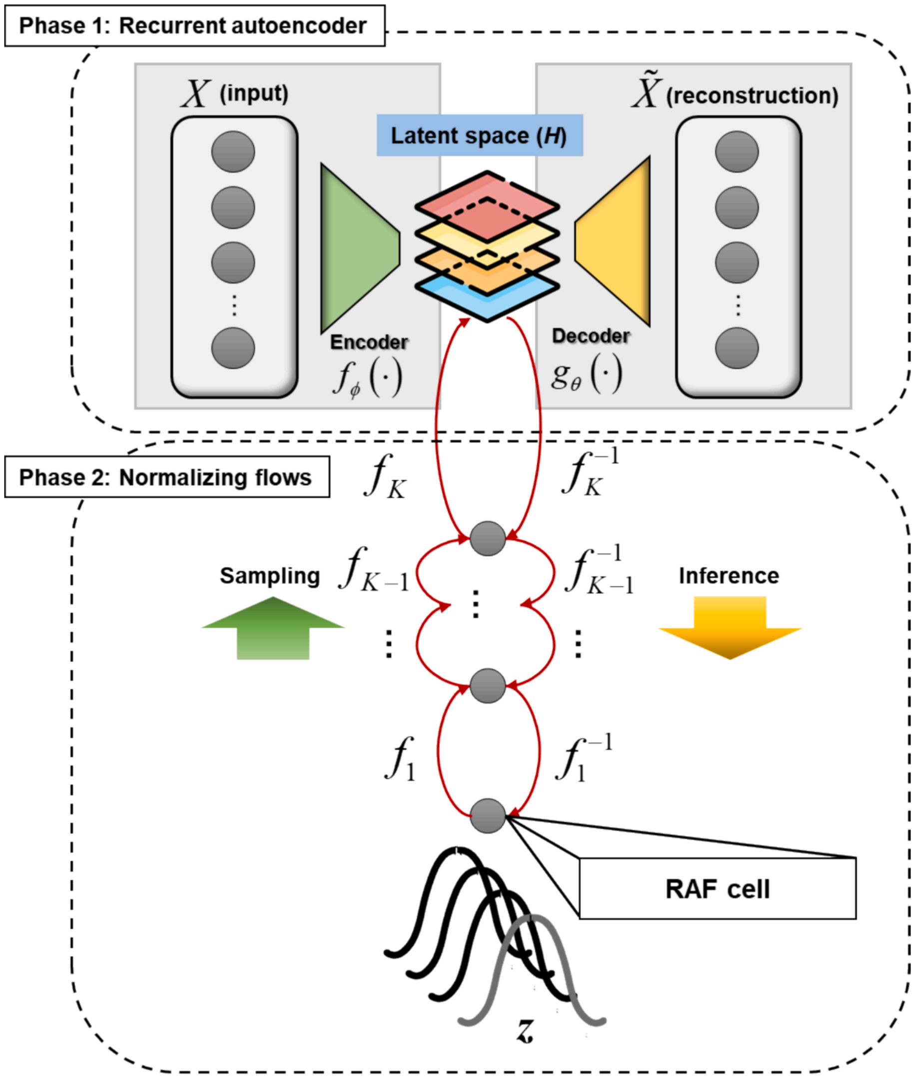


Fig. 5. Illustration of the proposed RAF-LSTM architecture.

- Epochs: 1000
- Optimizer: Adam; learning rate = 0.001
- Base distribution: StandardNormal(13)
- Hidden layer size in the GRU: 128
- Batch size: 64

3.2.2. Sensor fault detection

In a sensor validation framework, anomalies are detected in the sensor measurements based on metrics such as the squared prediction error (*SPE*) or Hotelling's T^2, which are computed from the process data transformation under normal conditions. This establishes a control limit for the density estimation with a certain confidence interval that can be used to determine whether an anomaly is occurring during the monitoring process. The statistics and control limits typically employed in fault detection are summarized in Section 2 of the SI (Han et al., 2018).

The fault detection method employed in the present study is a scaled anomaly score (Θ) with a range of [0, 1] computed as shown in Eq. (19):

$$\text{Anomaly score } (\Theta_i) = w_1 \cdot \breve{L}_{flow,i} + w_2 \cdot \breve{L}_{recons,i} \tag{19}$$

where w_1 and w_2 are weights used to balance the importance of each component ($w_1 = w_2 = 0.5$ in the present study), $\breve{L}_{flow}$ is the log-probability of the latent space, and $\breve{L}_{recons}$ is the reconstruction error obtained at instance i, computed as shown in Eqs. (20) and (21):

$$\breve{L}_{flow,i} = \log p_H(H)_i \tag{20}$$

$$\breve{L}_{recons,i} = \left| x_i - \tilde{x}_i \right| \tag{21}$$

The threshold for fault detection is established using Θ by taking the α quantile from its probability density, where $1 - \alpha$ is a confidence coefficient of 95 % ($\alpha = 0.05$), as computed in Eq. (22) (Papastefanopoulos et al., 2025):

$$\text{Threshold } (\tau) = percentile\left(\Theta_{i=1}^{N}, \alpha\right) \tag{22}$$

When monitoring the seal temperatures, anomaly scores are considered to represent faulty measurements when they exceed the established control limit τ; otherwise, they are considered normal.

Confusion matrices are utilized to evaluate the detection performance of the tested models. In general, fault detection methods adopt similar approaches in terms of the control charts and thresholds used to establish whether a measurement is faulty or not. A binary variable (0 for a normal measurement and 1 for a faulty measurement) is thus used to compare the results to the ground truth to assess detection performance. The selected metrics are described in Eqs. (23)–(26):

$$\text{Recall} = \frac{TP}{TP + FN} \tag{23}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{24}$$

$$\text{F1 - Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{25}$$

$$\text{False Positive Rate } (FPR) = \frac{FP}{FP + TN} \tag{26}$$

Recall measures the detection rate, which indicates how many of the actual faulty values are predicted correctly and is measured using the number of true positive (*TP*) and false negative (*FN*) cases. Precision indicates how many of the predicted faulty measurements are actually faulty using the number of *TP* and false positive (*FP*) cases. The F1 score represents the harmonic mean of precision and recall, while the false positive rate (*FPR*) indicates the proportion of normal measurements that are classified as faulty.

In addition to these threshold-based metrics, receiver operating characteristic (ROC) curves were constructed using the SPE statistic as the anomaly score. This analysis characterizes the trade-off between the true positive rate (TPR) and FPR across all possible thresholds, thereby providing a threshold-independent evaluation of model performance. The area under the ROC curve (AUC) is a metric that reflects the discrimination between normal and faulty measurements with a strong capability when this value is close to 1 (Lavazza et al., 2025).

3.2.3. Sensor fault reconstruction

Once a fault is detected, the reconstruction process aims to transform the faulty data based on the learned behavior under normal operating conditions in the latent or feature spaces (Yi et al., 2017). The reconstructed measurements are then compared with actual data under normal conditions. The performance metrics are the root mean squared error (RMSE) and the mean absolute percentage error (MAPE) (Eqs. (27) and (28), respectively) (Liu et al., 2018):

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(x_i - \tilde{x}_i\right)^2} \tag{27}$$

$$MAPE = \frac{100\%}{m}\sum_{i=1}^{m}\left|\frac{x_i - \tilde{x}_i}{x_i}\right| \tag{28}$$

where x is the observed TBM-EPB data under normal conditions, and $\tilde{x}$ is the reconstructed data.

3.3. Proposed method

The sensor validation approach proposed in the present study consists of offline modeling and online monitoring. The offline modeling process trains the models with normal TBM-EPB process data so that normal behavior is learned, while the online monitoring process validates the trained models by propagating the test data and detecting and reconstructing faulty measurements (Fig. 6). These two processes consist of the following steps:

- *Offline modeling:*

Step 1: Collecting training data under normal conditions.

Step 2: Standardizing the data to a zero mean and unit variance.

Step 3: Training state-of-the-art feature extractors, recurrent AEs, and NFs in the latent space.

Step 4: Computing the SPE and anomaly scores for the statistical method and the proposed RAF-LSTM method, respectively.

Step 5: Defining the control limits using density estimation for the statistical methods and taking the quantile from the anomaly score

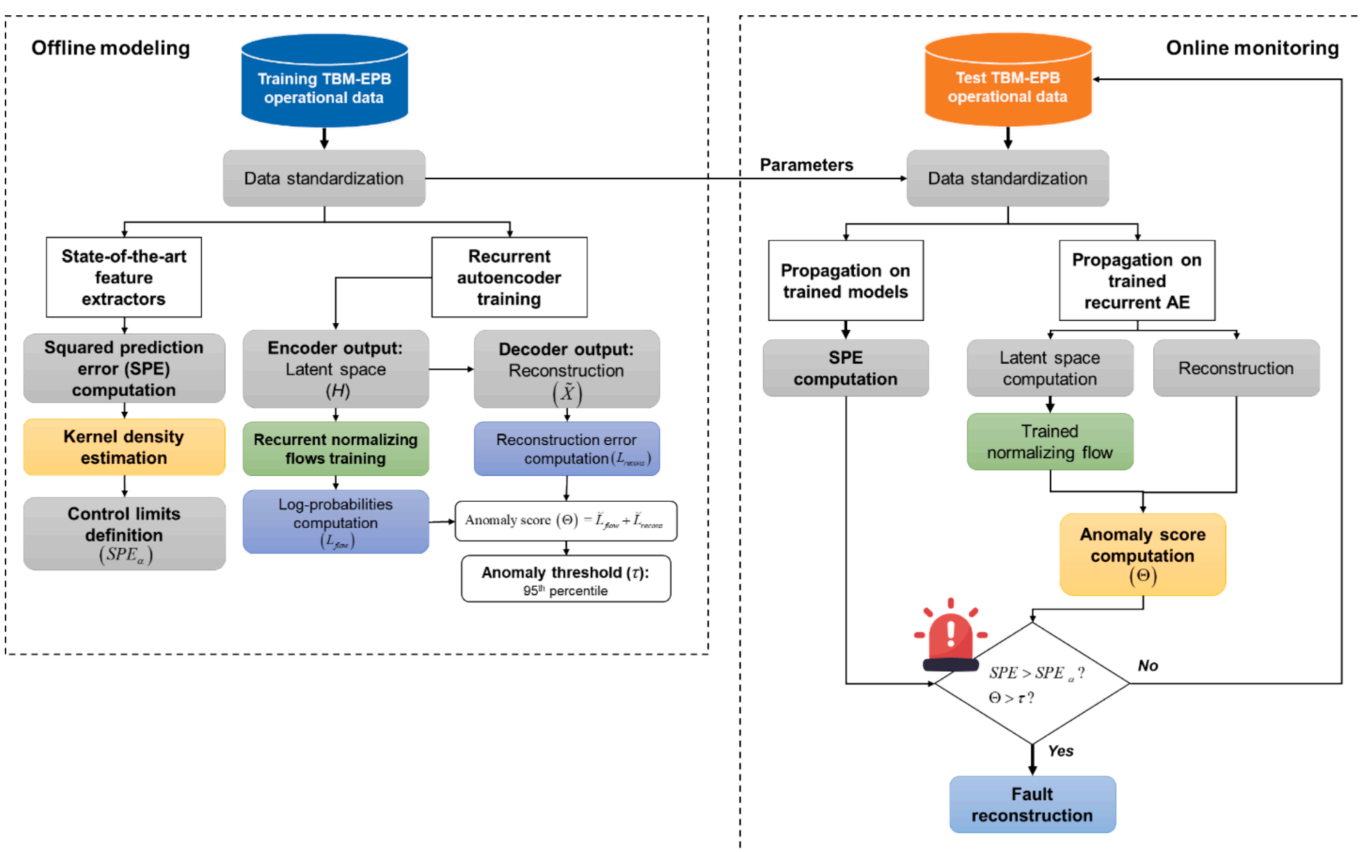


Fig. 6. Research framework for seal temperature sensor monitoring during TBM-EPB operation including offline modeling and online monitoring.

density for RAF-LSTM. These limits are calculated with a significance level of 0.05.

- *Online monitoring:*

Step 1: Collecting artificial faulty data and standardizing it with the training set parameters.

Step 2: Propagating the test data through the trained models, computing the latent space, and reconstructing the data.

Step 3: Computing the SPE and the anomaly scores and comparing them with the control limits to determine whether the measurements are faulty.

Step 4: Evaluating the performance of the detection and reconstruction of the faulty measurements.

RAF-LSTM is compared with state-of-the-art methods, including PCA, independent component analysis (ICA), KPCA, and neural methods, such as shallow AE and stacked AE (SAE), a variational AE with convolutional layers (VAE-CNN), and a masked autoregressive flow AE (MAF-AE). The dimensionality of the feature space and latent representations is set to the number of principal components that capture 85 % of the explained variance (i.e., $n = 13$) as explained in Section 2 of the SI (Du et al., 2020; Navi et al., 2018b).

The developed models are implemented in MATLAB R2023a and Python 3.8 with Keras, a deep learning package utilized in AE modeling, and PyTorch for NF training. The computer system has the following specifications: 13th Gen Intel® Core ™ i9-13900 K, 3000 MHz, 32.0 GB RAM, and a x64-based PC.

4. Results and discussion

4.1. Preliminary TBM-EPB data analysis

The proposed monitoring system for seal temperature sensors during TBM-EPB operation integrates recurrent neural structures, including AE and NFs. Because these models are designed to manage non-linear relationships and temporal dependencies in the input data, a preliminary analysis is conducted to assess their utility for the established set-up. The Hurst exponent is determined according to the rescaled range analysis (R/S) for the time interval and indicates the extent to which the time series depends on its past values (Lotfalinezhad and Maleki, 2020).

Figs. 7. (a)–(d) depict the relationship between R/S and the time interval for the four seal temperature sensors and their respective Hurst exponents. The red dashed lines ($H = 0.5$) represent Brownian motion with a normal distribution and no long memory, the purple dots are the regression results for the R/S analysis, and the continuous light blue line is the corresponding Hurst line. The generated Hurst lines are positioned below the normal line with a smaller slope. The temperature time series can thus be classified as anti-persistent ($H < 0.5$). This means that, when the seal temperature exhibits an increase at time t, there is a high probability that it will be followed by a decrease at $t + 1$; similarly, a

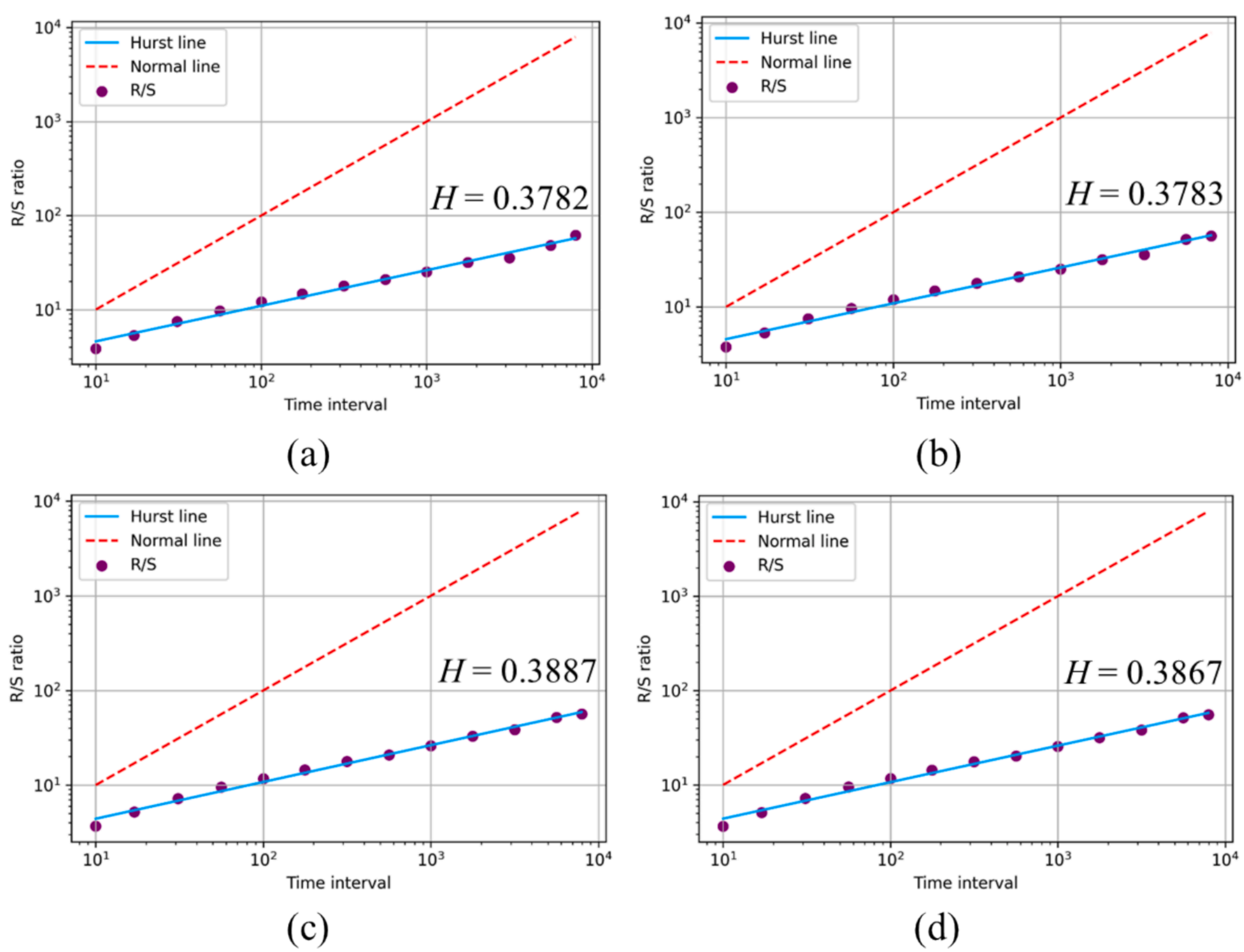


Fig. 7. R/S analysis and Hurst exponents for the seal temperature sensors (a) ST1, (b) ST2, (c) ST3, and (d) ST4.

decrease at time t is likely to be followed by an increase at $t + 1$, leading to a wave-like pattern. Therefore, because the analyzed time series does not follow normality, non-linear relationships and temporal dependencies can be better processed by the advanced neural models, especially those with recurrent characteristics because they utilize past information.

The anti-persistent behavior observed in the seal temperature series implies that increases are likely followed by decreases and vice-versa, resulting in oscillatory short-term reversals. These dynamics are not well captured by linear monitoring approaches, e.g., PCA, but they align with the use of recurrent neural structures that can retain and process temporal dependencies. For these reasons, LSTM memory cells were selected for the AE reconstruction, and GRU-based RAF were employed for density estimation.

4.2. Seal temperature sensor monitoring during TBM-EPB operation

The proposed RAF-LSTM method consists of a recurrent AE that generates a latent space and a reconstruction, after which recurrent NFs learn the probability distribution over this latent space. The reconstruction error and the log-probability of this structure generate anomaly scores that define the control limit using a 95 % confidence interval.

4.2.1. Fault detection for the seal temperature sensors

To detect faulty measurements, RAF cells are used to estimate the density of the latent space of the recurrent AE, utilizing LSTM as the memory cell. After training both the AE and NFs, anomaly score Θ is computed for each instance utilizing the log-probabilities and the reconstruction error, computing a threshold τ based on a 95 % confidence interval. Fig. 8 presents the probability density for the anomaly scores from the training set, which contains data collected under normal conditions, and the test set, which contains measurements with faulty data.

The detection performance of the proposed method is assessed by propagating the test data with artificial faults and comparing the output with previously reported state-of-the-art and neural methods. The methods employed for the comparison compute the SPE, and their limit is determined using a kernel density estimator with a 95 % confidence interval. The SPE control chart for these methods is presented in Section 3 of the SI. The detection results for the models are binarized and investigated as a classification problem.

Fig. 9 presents the confusion matrices for the compared methods, and Table 4 summarizes the corresponding performance metrics for fault detection. The confusion matrices depict the number and percentage of TP, FP, FN, and TN cases for two classes (*Fault* and *No Fault*). Figs. 9 (a)–(c) present the detection results for the state-of-the-art methods. PCA has an FP rate of 42.2 % and an FN rate of 11.8 %, representing errors in correctly classifying sensor faults. In contrast, the TP and TN rates are 26.2 % and 19.9 %, respectively, indicating the correct classification of faults. This leads to an accuracy of 46.03 %, a precision of 38.29 %, a recall of 68.92 %, and an F1-score of 0.49. Overall, type I and type II errors account for more than 40 % of the cases for KPCA and more than 50 % for the PCA and ICA methods, with high rates of FPs that classify faulty conditions as non-faulty (i.e., type I errors). These high error rates result from their reliance on linear projections or independent components without modeling sequential dependencies.

The neural methods exhibit a strong performance when detecting faults in the temperature sensors. Both AE and SAE models achieve recalls above 90 %, although they misclassify approximately 42 % of normal cases as faulty (Type I errors). The VAE-CNN model further improves recall to 100 % with moderate precision (46.2 %). Moreover, the flow-based MAF-AE also demonstrate competitive performance but with variable accuracy across sensors. In constrast, the proposed RAF-LSTM achieved outstanding results with an accuracy of 96.6 %, precision of 91.9 %, recall of 100 %, and an F1-score of 0.96, reducing type I errors to 3.6 % and eliminating type II errors entirely.

While confusion matrices rely on a fixed threshold, the ROC analysis provides a threshold-independent view of the discriminative

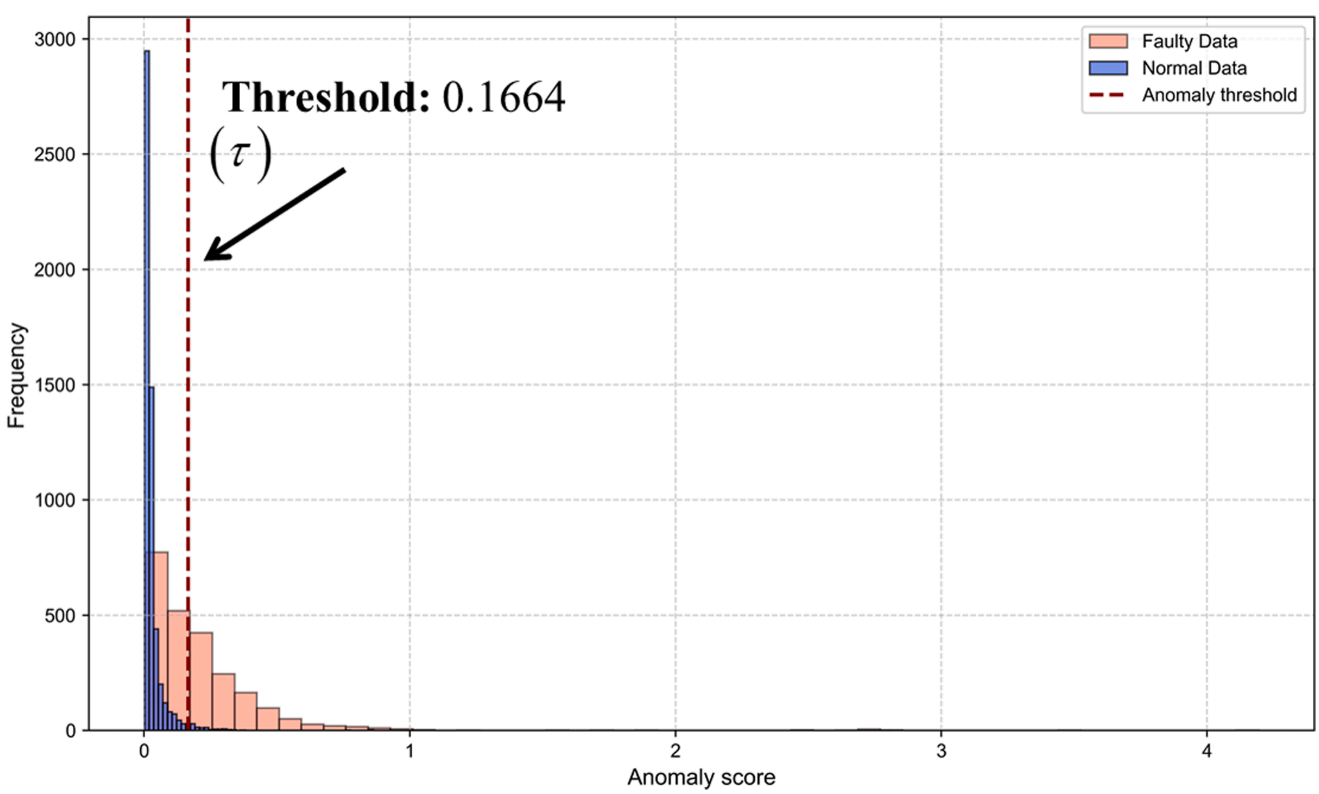


Fig. 8. Distribution of anomaly scores computed for normal and faulty data with threshold τ using a confidence interval of 95%.

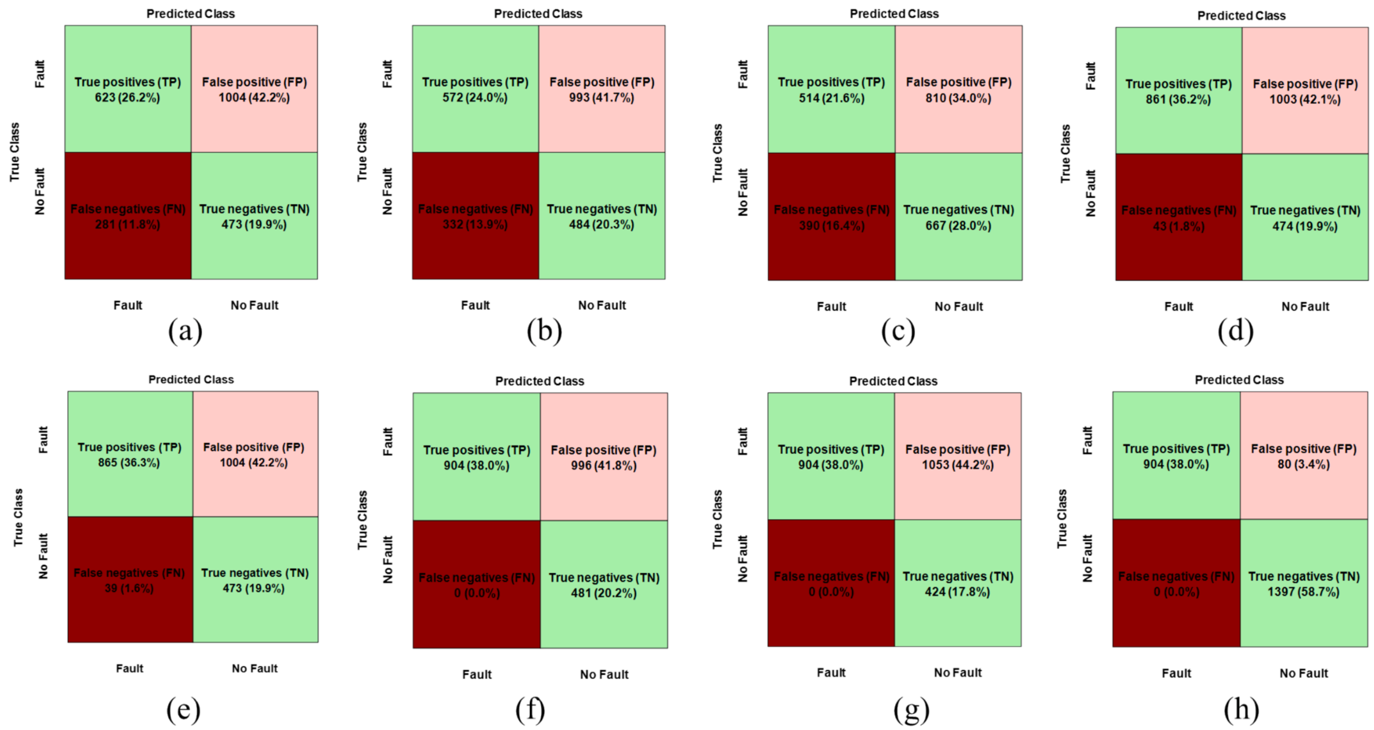


Fig. 9. Confusion matrices for seal temperature sensor fault detection using (a) PCA, (b) ICA, (c) KPCA, (d) AE, (e) SAE, (f) MAF-AE, (g) VAE-CNN, and (h) RAF-LSTM methods.

Table 4
Detection performance metrics for seal temperature sensor fault across different methods. The proposed RAF-LSTM significantly outperforms all baselines with 96.6% accuracy, 91.9% precision, and 100% recall, with the highest F1-score (0.96).

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
PCA	46.03	38.29	68.92	0.49
ICA	44.35	36.55	63.27	0.46
KPCA	49.60	38.82	58.86	0.46
AE	56.07	46.19	95.24	0.62
SAE	56.19	46.28	95.69	0.62
MAF-AE	49.10	40.86	76.11	0.53
VAE-CNN	55.77	46.19	100.00	0.63
RAF-LSTM	**96.64**	**91.87**	**100.00**	**0.96**

performance of the compared methods. ROC curves were constructed utilizing the SPE statistic as an anomaly score, as illustrated in Fig. 10. The state-of-the-art linear methods show poor discriminative performance, with AUC values of 0.46 (PCA), 0.49 (ICA), and 0.54 (KPCA). Moreover, neural baselines provide moderate improvements, with AUC values of 0.63 (AE, SAE), 0.55 (VAE-CNN), and 0.56 (MAF-AE). In contrast, the proposed RAF-LSTM achieved an AUC of 0.94, indicating superior discrimination between normal and faulty sensor states.

Overall, the proposed RAF-LSTM method exhibits superior detection performance compared to both traditional statistical methods and advanced neural or flow-based approaches. Its structure integrates recurrence with flow-based density estimation, enabling the joint modeling of temporal dependencies and probabilistic scoring. AEs only reconstruct individual data points, neglecting long-term temporal patterns, while RAF-LSTM is capable of both density estimation due to the RAF and accurate reconstruction via LSTM, generating a reliable anomaly score that describes process behavior under normal and faulty conditions.

4.2.2. Reconstruction of faulty temperature sensors

Once a fault is detected, reconstruction needs to be conducted to force the faulty measurements to follow normal conditions. In the proposed method, this reconstruction is solely conducted through the AE with LSTM cells. Table 5 presents the performance metrics for the fault reconstruction using the tested methods.

Among the statistical approaches, PCA and ICA exhibit the largest errors, with MAPEs ranging from 7.4 % to 53.5 %, because they assume that the process data can be explained in a linear subspace, in the case of the PCA, and statistical independence for ICA, there is a limitation with their ability to capture complex, nonlinear, and dependent dynamics of TBM-EPB operational data. KPCA introduces nonlinearity through kernel functions and achieves improved accuracy (MAPE: 6.1–13.9 %), but still does not incorporate temporal dependencies, restricting its performance in sequential reconstruction.

Neural methods generally deliver stronger reconstruction performance compared with statistical baselines. The shallow AE model reduces reconstruction errors to MAPEs between 7.7 % and 20.1 % by learning nonlinear mapping between input and latent representations. The SAE model further improves robustness (MAPE: 7.4 – 19.3 %) through its deeper layered structure, which captures more complex hierarchical features. However, both AE and SAE remain limited in their ability to capture sequential dependencies, as their reconstruction is based on independent input windows rather than temporal dynamics.

Advanced neural approaches exhibit additional gains. For instance, the VAE-CNN combines the generative capacity of VAE with convolutional feature extraction, enabling it to capture both global variability and local temporal patterns in the sensor data. This leads to stable improvements (MAPE: 6.2 – 14.9 %) across all sensors, outperforming traditional AE in several cases. On the other hand, the MAF-AE approach integrates two models that enhance density estimation in the latent space of the AE, allowing probabilistic scoring of reconstructions. Although this approach yields competitive results (MAPE: 8.6 – 22.9 %), its performance varies depending on the sensor, possibly to sensitivity in learning long-range temporal correlations, as its mechanism does not digest sequential data.

In contrast, the proposed method consists of an LSTM-based AE that not only conducts its transformations through non-linear activations but

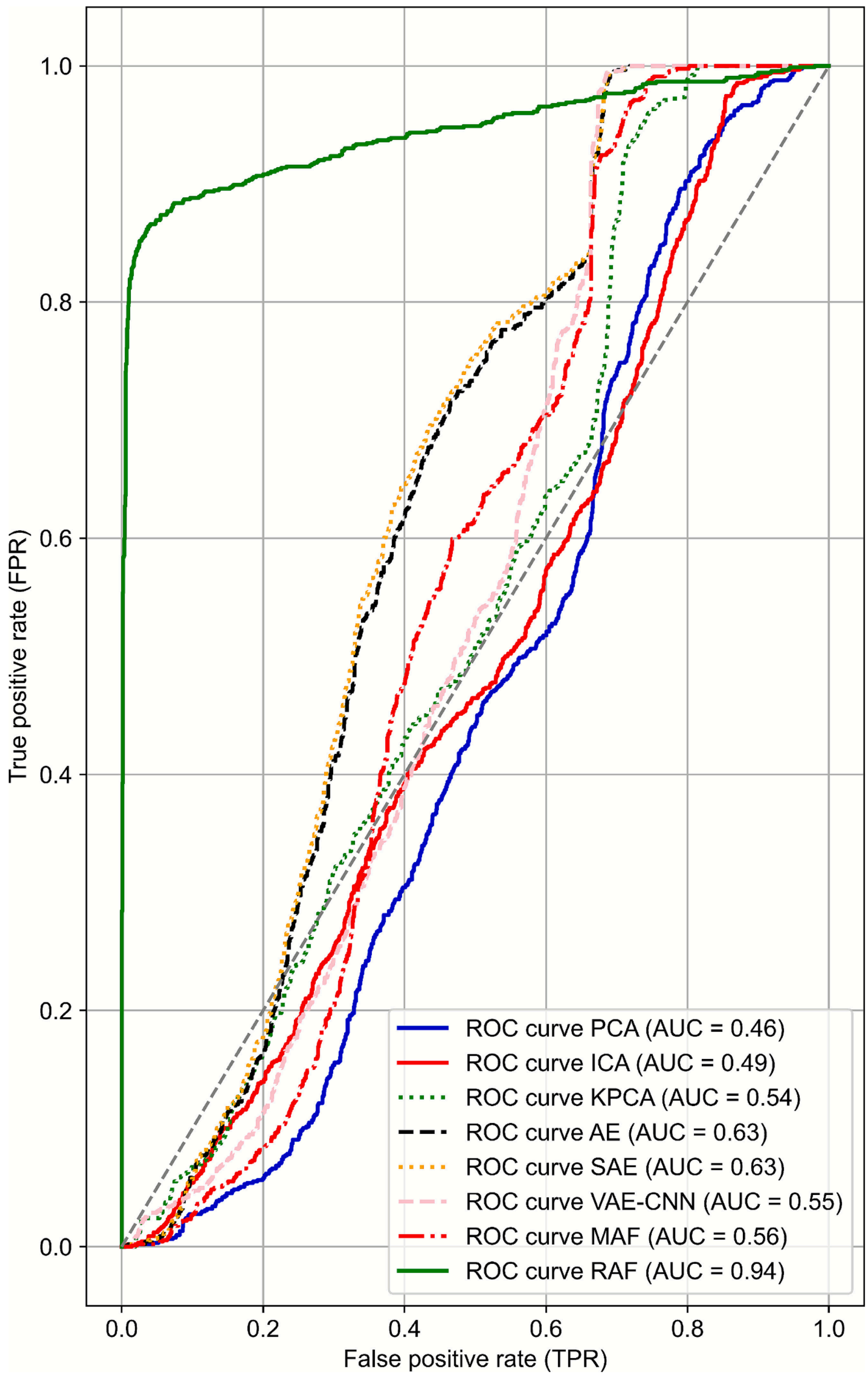


Fig. 10. ROC curves and AUC metrics for seal temperature sensor fault detection. RAF-LSTM achieves AUC of 0.94, substantially outperforming statistical and neural baselines (AUC < 0.65).

also incorporates memory cells to process sequential data from the TBM-EPB. This leads to a superior performance for the RAF-LSTM, with a MAPE of 1.14–1.62 % and an RMSE of 0.51–0.57C. Fig. 11 displays the reconstruction results for the LSTM-based AE, including a comparison between the actual data and the coefficient of determination (R^2), which is more than 0.9 for all of the reconstructed sensors in faulty conditions. The proposed method can also identify the occurrence of peaks via sequence modeling.

4.3. Operational implications for TBM-EPB excavation

This study developed a sensor validation approach for the monitoring of seal temperature sensors during TBM-EPB tunneling by coupling feature extractors in a memory-cell AE with a density estimator using recurrent NFs for detecting and reconstructing faulty sensors. Recurrence-based LSTM within the AE and GRU cells in the NF were used due to the anti-persistent trend in the measurement data recorded during shield machine operation. As a result, the proposed RAF-LSTM method outperformed state-of-the-art and neural methods that extract

Table 5
Performance metrics for seal temperature sensor reconstruction using different methods. The proposed RAF-LSTM achieves the lowest errors (RMSE = 0.51 – 0.57, MAPE = 1.1 – 1.6 %).

Method	Sensor	RMSE (C)	MAPE (%)
PCA	ST1	18.45	53.49
	ST2	12.86	37.15
	ST3	3.56	9.93
	ST4	3.43	7.42
ICA	ST1	17.57	49.23
	ST2	10.79	28.44
	ST3	5.91	20.07
	ST4	5.37	16.37
KPCA	ST1	5.35	13.88
	ST2	2.56	6.08
	ST3	2.23	6.14
	ST4	2.75	7.17
AE	ST1	3.01	7.73
	ST2	4.71	14.38
	ST3	5.95	20.08
	ST4	2.93	8.42
SAE	ST1	3.42	8.72
	ST2	4.16	12.60
	ST3	5.69	19.33
	ST4	2.68	7.42
MAF-AE	ST1	7.83	22.92
	ST2	4.81	15.05
	ST3	5.57	18.80
	ST4	3.02	8.59
VAE-CNN	ST1	4.17	11.85
	ST2	4.10	12.37
	ST3	4.58	14.89
	ST4	2.42	6.21
RAF-LSTM	ST1	**0.51**	**1.14**
	ST2	**0.57**	**1.38**
	ST3	**0.57**	**1.62**
	ST4	**0.51**	**1.21**

latent representations from process data through linear, kernel-based, and non-linear transformations but neglect time-dependent features.

Given the operating conditions for TBM-EPB machines, the following implications can be drawn from the results of the present study:

- *System failure prevention*: Because abnormal thermal behavior may indicate lubrication failure or seal damage, the proposed robust approach for temperature sensor monitoring that detects and reconstructs failures can output accurate information on the state of the seals, preventing catastrophic seal failure and avoiding unnecessary downtime and repair costs. It should be noted that abnormal seal temperature readings may not always serve as an early-warning indicator, as these anomalies can occur once substantial wear or lubrication breakdown have already developed. In tunneling, upstream parameters such as cutter torque, motor current, thrust force, chamber pressure, or seal pressure can provide earlier indications of cutter seal distress. While the present work focused on detecting and reconstructing faulty temperature sensors, the proposed framework is not limited and could be extended to integrate additional variables. In this context, a multi-sensor monitoring strategy would strengthen early detection capabilities and enhance the operational payoff of predictive maintenance in TBM-EPB projects.
- *Work safety*: Undetected sensor faults may disguise overheating, which accelerates seal degradation and allows the ingress of soil or pressurized slurry into the drive. For these reasons, the proposed fault-tolerant sensor monitoring system has the potential to ensure the integrity of the tunnel face and the shield while mitigating risks to worker safety.
- *Reducing false alarms and operational downtime*: False alarms in a sensor monitoring system can prompt emergency stops, making it difficult to meet deadlines and increasing operational costs. Therefore, a reliable validation system that minimizes false negatives is desirable to ensure productive tunneling operations.

In addition to these implications, tunnel engineering research is currently investigating the development of digital twins and model-based control strategies that rely on high-fidelity sensors; thus, the proposed monitoring framework can be extended for the reading of other sensors in TBM-EPB projects, maintaining reliable information even when the physical sensors are degraded. Furthermore, it is relevant to analyze the coupling between seal temperature and upstream operational parameters, such as cutter torque, motor current, thrust force, and chamber or seal pressures. These variables often interact with thermal behavior, and their joint analysis could enhance engineering interpretation. While this study focused on temperature sensor validation, the RAF-LSTM framework can be extended to incorporate such multivariate relationships in future applications.

From a deployment perspective, the computational requirements of the proposed framework are modest once training is completed, while the model inference for fault detection and reconstruction can be performed near real-time, making integration into TBM supervisory control and data acquisition (SCADA) systems feasible. Additionally, the probabilistic scoring by RAF aligns with digital twin concepts, where virtual replicas of machine behavior require reliable inputs for simulation and decision making. Embedding the proposed method into SCADA or digital twin environments would enable operators to continuously validate sensor health, reducing false alarms and supporting condition-based maintenance in practice.

5. Conclusions and future research

This study proposed a hybrid RAF-LSTM framework for cutter seal temperature sensor validation in EPB tunneling. By combining a recurrent AE for temporal reconstruction with a RAF for probabilistic scoring, the method was able to detect and reconstruct faulty sensor signals more effectively than statistical and neural baselines. A dataset consisting of 42 operational variables for four seal temperature sensors in a TBM-EPB project was considered, with the data split into training and test sets. The training set contained data collected under normal conditions, while the test set contained synthetic failures.

The fault detection performance of the proposed framework was assessed by computing the anomaly scores for the test set and classifying the measurements as faulty or non-faulty. The RAF-LSTM achieved a recall of 100 %, an accuracy of 96.6 %, no FN, and an FPR of only 3.6 %. In contrast, other state-of-the-art methods yielded accuracies below 60 % and considerably higher error rates, which could lead to unnecessary halts to TBM-EPB operation. ROC analysis further confirmed the robustness of the proposed method, with an AUC of 0.94 compared to values below 0.65 for all baselines, demonstrating its superior threshold-independent discriminative capability. For reconstruction, the proposed method consistently yielded the lowest errors, with RMSE values of 0.51 – 0.57C and MAPE values of 1.1 – 1.6 %, representing up to an 80 – 90 % reduction compared with the best-performing baselines. These results confirm that the incorporation of memory cells for temporal sequence modeling, combined with flow-based probabilistic scoring, enables RAF-LSTM to robustly detect and accurately reconstruct faulty temperature sensor signals even under harsh operational conditions.

A limitation of the present study is that it was based on one tunneling

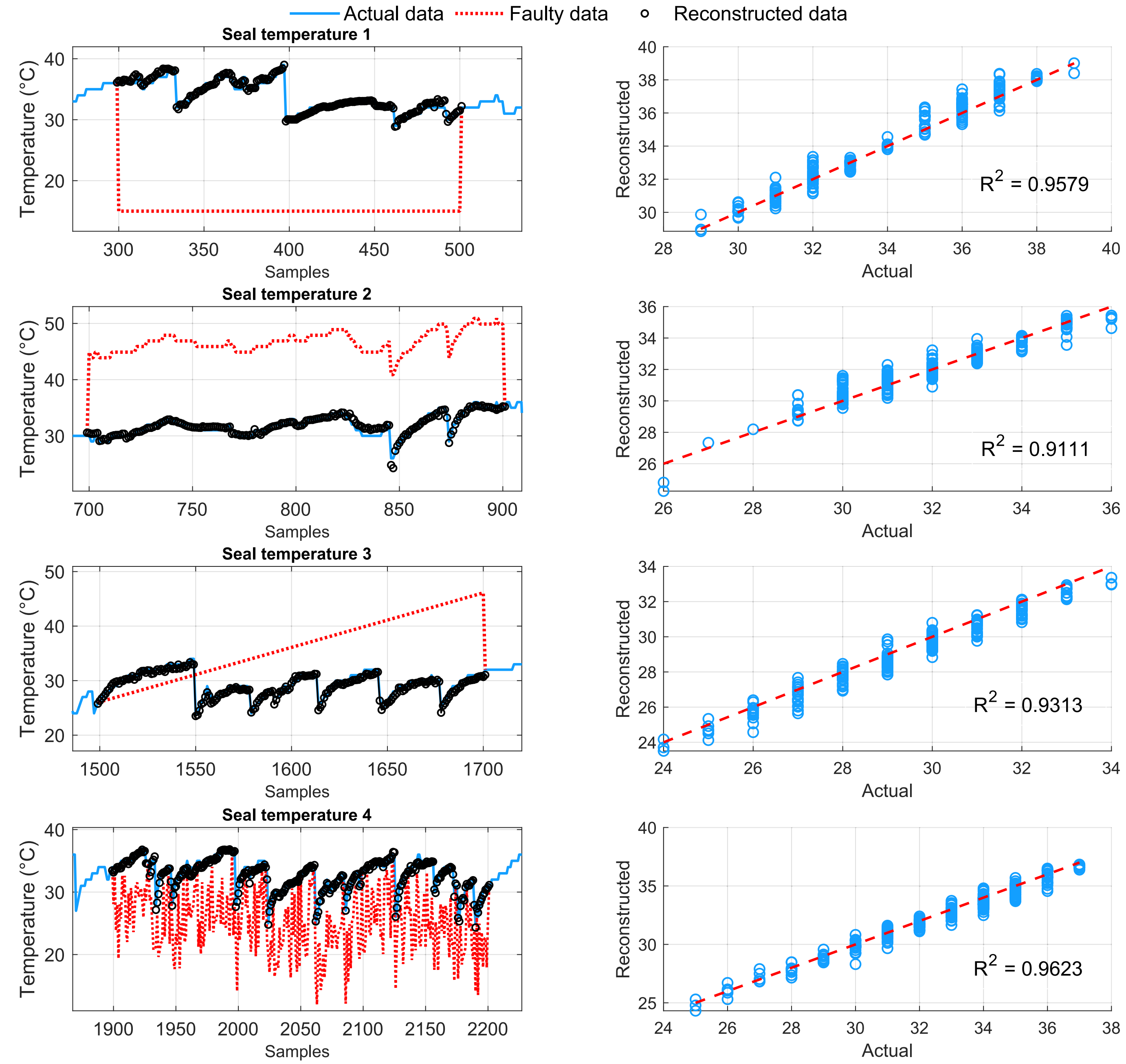


Fig. 11. Reconstruction results for the faulty seal temperature sensors using the proposed RAF-LSTM method, including a time series comparison (left) and scatter plots of actual and reconstructed data with the coefficient of determination (right).

project with four seal temperature sensors. Although this provided a focused case study to validate the proposed framework, further cross-project and cross-section testing is necessary to assess robustness under varying geological and operational conditions. Considering that the proposed RAF-LSTM method learns temporal dynamics directly from the data without relying on project-specific assumptions, it is well-suited for extension to larger sensor networks and multiple tunneling projects in future studies.

CRediT authorship contribution statement

Jorge Loy-Benitez: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Je-Kyum Lee:** Visualization. **Myung Kyu Song:** Writing – review & editing, Formal analysis, Data curation. **Fabian Cabrera Guerra:** Visualization. **Sean Seungwon Lee:** Supervision, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the BK21 FOUR (Fostering Outstanding Universities for Research) program (No. 4299990314624) through the National Research Foundation of Korea (NRF) funded by the Korea Ministry of Education and the Korea Institute for Advancement of

Technology (KIAT) grant funded by the Korea Government (MOTIE) (RS-2024-00410787, HDR Program for Industrial Innovation).

Appendix A

Table A1
Operational variables of the case study EPB and descriptive statistics.

Feature	Unit	Mean	Standard deviation	Minimum	Median	Maximum
NetStroke	mm	630.82	369.91	0.00	640.00	1361.00
Top Earth Pressure	kPa	16.03	25.74	0.00	0.00	169.00
Left Earth Pressure	kPa	48.14	42.48	0.00	34.00	249.00
Right Earth Pressure	kPa	49.23	40.55	0.00	37.00	220.00
Bottom Earth Pressure	kPa	104.09	39.19	2.00	98.50	271.00
Grout Volume (Ring)	m^3	2.15	2.46	0.00	1.65	36.22
Screw Muck Volume (Ring)	m^3	32.46	30.85	0.00	29.21	703.53
Bulk Analyzer Volume (Ring)	m^3	32.42	26.95	0.00	28.08	234.83
Polymer Volume (Ring)	m^3	6.87	13.95	0.00	3.84	152.52
Gate stroke	%	35.06	22.08	0.00	32.00	98.00
Jack Speed average	mm/min	13.73	5.29	0.00	14.00	32.00
Cutter RPM	min^{-1}	1.88	0.18	0.00	1.82	2.47
Pitching	Degrees	0.22	0.40	−1.33	0.29	0.88
Rolling	Degrees	0.01	0.32	−1.16	0.00	1.85
Net Excavation Time	min	62.90	56.51	0.00	52.40	529.00
Screw Rotation speed	min^{-1}	2.07	1.10	0.00	2.10	8.00
Screw Pressure	MPa	3.91	3.09	0.00	2.80	21.40
Screw normal rotation	min^{-1}	0.92	0.27	0.00	1.00	1.00
Screw reverse rotation	min^{-1}	0.00	0.04	0.00	0.00	1.00
Screw Torque	kN-m	34.30	27.18	0.00	24.00	188.00
Cutter CCW	on/off	0.48	0.50	0.00	0.00	1.00
Cutter CW	on/off	0.51	0.50	0.00	1.00	1.00
Left SJ speed	mm/min	13.70	5.29	0.00	14.00	34.00
Right SJ speed	mm/min	13.76	5.40	0.00	14.00	34.00
Left SJ stroke	mm	1349.65	372.35	606.00	1349.00	2245.00
Right SJ stroke	mm	1337.54	371.65	561.00	1337.00	2210.00
Shield pressure	MPa	12.96	5.98	0.00	11.70	33.90
Total Cutter motor torque	%	38.88	18.06	0.00	41.30	85.00
Tail Seal Pressure	MPa	3.55	1.41	0.00	3.30	7.90
Tail Seal Tank Level	mm	324.89	219.63	6.00	317.00	907.00
Chainage (Head)	m	34938.31	90.36	34754.19	34960.28	35052.80
No.1 Cutter seal temperature	℃	30.54	5.15	15.00	31.00	47.00
No.2 Cutter seal temperature	℃	29.79	4.83	15.00	31.00	44.00
No.3 Cutter seal temperature	℃	29.87	4.62	15.00	31.00	42.00
No.4 Cutter seal temperature	℃	29.98	4.64	15.00	31.00	42.00
Gross Excavation Time	min	779.93	5729.24	0.00	96.15	171937.40
No.1 Cutter motor current	A	23.99	11.49	0.00	24.00	115.00
No.2 Cutter motor current	A	23.88	11.13	0.00	24.00	52.00
No.3 Cutter motor current	A	18.78	13.67	0.00	19.00	52.00
No.4 Cutter motor current	A	23.58	11.06	0.00	24.00	52.00
Thrust force	kN	15744.34	7576.11	0.00	13805.50	39916.00
Cutter Torque	kNm	4397.44	2041.95	0.00	4670.50	9614.00

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tust.2025.107264.

Data availability

Data will be made available on request.

References

Akhlaghi, M.A., Bagherpour, R., Hoseinie, S.H., 2025. Real-time monitoring of disc cutter wear in tunnel boring machines: a sound and vibration sensor-based approach with machine learning technique. J. Rock Mech. Geotech. Eng. 17, 1700–1722. https://doi.org/10.1016/j.jrmge.2024.04.030.

Ba-Alawi, A.H., Vilela, P., Loy-Benitez, J., Heo, S.K., Yoo, C.K., 2021. Intelligent sensor validation for sustainable influent quality monitoring in wastewater treatment plants using stacked denoising autoencoders. J. Water Process Eng. 43, 102206. https://doi.org/10.1016/j.jwpe.2021.102206.

Brooks, S., K., Bauer, M., 2018. Sensor validation and reconstruction: experiences with commercial technology. Control Eng. Pract. 77, 28–40. https://doi.org/10.1016/J.CONENGPRAC.2018.04.003.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf. 1724–1734. doi: 10.3115/v1/d14-1179.

Cortez, B., Carrera, B., Kim, Y.J., Jung, J.Y., 2018. An architecture for emergency event prediction using LSTM recurrent neural networks. Expert Syst. Appl. 97, 315–324. https://doi.org/10.1016/j.eswa.2017.12.037.

Dinh, L., Sohl-Dickstein, J., Bengio, S., 2017. Density estimation using real NVP. 5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.

Du, W., Zhang, Y., Zhou, W., 2020. Modified non-Gaussian multivariate statistical process monitoring based on the Gaussian distribution transformation. J. Process Control 85, 1–14. https://doi.org/10.1016/J.JPROCONT.2017.12.001.

Durkan, C., Besakov, A., Murray, I., Papamakarios, G., 2020. nflows: normalizing flows in PyTorch. doi: 10.5281/zenodo.4296287.

Elbaz, K., Shen, S.L., Zhou, A., Yin, Z.Y., Lyu, H.M., 2021. Prediction of disc cutter life during shield tunneling with AI via the incorporation of a genetic algorithm into a GMDH-Type neural network. Engineering 7, 238–251. https://doi.org/10.1016/j.eng.2020.02.016.

Fu, J., Wu, D., Lan, H., Ji, Z., Li, W., Xia, Y., 2021. Online monitoring and analysis of TBM cutter temperature: a case study in China. Meas. J. Int. Meas. Confed. 174, 109034. https://doi.org/10.1016/j.measurement.2021.109034.

Gomaa, M., Schade, S., Bao, D.W., Xie, Y.M., 2023. Automation in rammed earth construction for industry 4.0: Precedent work, current progress and future prospect. J. Clean. Prod. 398, 136569. https://doi.org/10.1016/j.jclepro.2023.136569.

Han, H.G., Dong, L.X., Qiao, J.F., 2021. Data-knowledge-driven diagnosis method for sludge bulking of wastewater treatment process. J. Process Control 98, 106–115. https://doi.org/10.1016/j.jprocont.2021.01.001.

Han, H.-G., Liu, Z., Guo, Y.-N., Qiao, J.-F., 2018. An intelligent detection method for bulking sludge of wastewater treatment process. J. Process Control 68, 118–128. https://doi.org/10.1016/j.jprocont.2018.05.002.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Kalayci Sahinoglu, U., Ozer, U., 2020. The prediction of cutter wear from temperature measurements on TBM discs and cutting face. Arab. J. Geosci. 13. https://doi.org/10.1007/s12517-020-5188-0.

Kang, X., Li, Y., Zhang, Y., Ma, N., Wen, L., 2024. Anomaly detection in concrete dam using memory-augmented autoencoder and generative adversarial network (MemAE-GAN). Autom. Constr. 168, 105794. https://doi.org/10.1016/j.autcon.2024.105794.

Kobyzev, I., Prince, S.J.D., Brubaker, M.A., 2021. Normalizing Flows: an Introduction and Review of Current Methods. IEEE Trans. Pattern Anal. Mach. Intell. 43, 3964–3979. https://doi.org/10.1109/TPAMI.2020.2992934.

Lavazza, L., Morasca, S., Rotoloni, G., 2025. Software defect Prediction evaluation: New metrics based on the ROC curve. Inf. Softw. Technol. 187. https://doi.org/10.1016/j.infsof.2025.107865.

Li, L., Tao, J.F., Yu, H.D., Huang, Y.X., Liu, C.L., 2017. Online condition monitoring of gripper cylinder in TBM based on EMD method. Chinese J. Mech. Eng. (english Ed.) 30, 1325–1337. https://doi.org/10.1007/s10033-017-0187-0.

Liu, H., Kim, M., Kang, O., Sankararao, B., Kim, J., Kim, J.C., Yoo, C.K., 2012. Sensor validation for monitoring indoor air quality in a subway station. Indoor Built Environ. 21, 205–211. https://doi.org/10.1177/1420326X11419342.

Liu, H., Kang, O., Kim, M., Oh, T., Lee, S., Kim, J.T., Yoo, C., 2013. Sustainable monitoring of indoor air pollutants in an underground subway environment using self-validating soft sensors. Indoor Built Environ. 22, 94–109. https://doi.org/10.1177/1420326X12469744.

Liu, C., Wang, Z., Liu, H., Cui, J., Huang, X., Ma, L., Zheng, S., 2024. Prediction of surface settlement caused by synchronous grouting during shield tunneling in coarse-grained soils: a combined FEM and machine learning approach. Undergr. Sp. 16, 206–223. https://doi.org/10.1016/j.undsp.2023.10.001.

Liu, H., Yang, C., Huang, M., Wang, D., Yoo, C.K., 2018. Modeling of subway indoor air quality using Gaussian process regression. J. Hazard. Mater. 359, 266–273. https://doi.org/10.1016/j.jhazmat.2018.07.034.

Lotfalinezhad, H., Maleki, A., 2020. TTA, a new approach to estimate Hurst exponent with less estimation error and computational time. Phys. A Stat. Mech. Its Appl. 124093. https://doi.org/10.1016/J.PHYSA.2019.124093.

Loy-Benitez, J., Heo, S., Yoo, C., 2020. Soft sensor validation for monitoring and resilient control of sequential subway indoor air quality through memory-gated recurrent neural networks-based autoencoders. Control Eng. Pract. 97, 104330. https://doi.org/10.1016/J.CONENGPRAC.2020.104330.

Loy-Benitez, J., Lee, H.K., Song, M.K., Lee, J.K., Lee, S.S., 2024. Real-time unsupervised monitoring of earth pressure balance shield-induced sinkholes in mixed-face ground conditions via convolutional variational autoencoders. Tunn. Undergr. Sp. Technol. 152, 105908. https://doi.org/10.1016/j.tust.2024.105908.

Mern, J., Morales, P., Kochenderfer, M.J., 2020. Towards Recurrent Autoregressive Flow Models.

Navi, M., Meskin, N., Davoodi, M., 2018. Sensor fault detection and isolation of an industrial gas turbine using partial adaptive KPCA. J. Process Control 64, 37–48. https://doi.org/10.1016/J.JPROCONT.2018.02.002.

Pan, Y., Fu, X., Zhang, L., 2022. Data-driven multi-output prediction for TBM performance during tunnel excavation: an attention-based graph convolutional network approach. Autom. Constr. 141, 104386. https://doi.org/10.1016/j.autcon.2022.104386.

Pang, G., Shen, C., Cao, L., Hengel, A., Van Den, 2021. Deep learning for anomaly detection: a review. ACM Comput. Surv. 54. https://doi.org/10.1145/3439950.

Papastefanopoulos, V., Linardatos, P., Kotsiantis, S., 2025. Combining normalizing flows with decision trees for interpretable unsupervised outlier detection. Eng. Appl. Artif. Intel. 141, 109770. https://doi.org/10.1016/j.engappai.2024.109770.

Qin, C.J., Wu, R.H., Huang, G.Q., Tao, J.F., Liu, C.L., 2023. A novel LSTM-autoencoder and enhanced transformer-based detection method for shield machine cutterhead clogging. Sci. China Technol. Sci. 66, 512–527. https://doi.org/10.1007/s11431-022-2218-9.

SKF, 2019. Bearing and seal design considerations for tunnel boring machines [WWW Document]. URL https://www.skf.com/group/news-and-events/news/2019/2019-01-01-bearing-and-seal-design-considerations-for-tunnel-boring-machines.

Tariq, S., Loy-Benitez, J., Yoo, C.K., 2024. Multi-sensor fault detection and correction for automated IAQ monitoring in smart buildings through attention-aware autoencoders with spatial prediction module. J. Build. Eng. 96, 110573. https://doi.org/10.1016/j.jobe.2024.110573.

Wang, S., Chen, Y., 2004. Sensor validation and reconstruction for building central chilling systems based on principal component analysis. Energy Convers. Manag. 45, 673–695. https://doi.org/10.1016/S0196-8904(03)00180-8.

Wang, L., Sun, W., Long, Y., Yang, X., 2018. Reliability-based performance optimization of tunnel boring machine considering geological uncertainties. IEEE Access 6, 19086–19098. https://doi.org/10.1109/ACCESS.2018.2821190.

Wang, S., Xing, J., Jiang, Z., Dai, Y., 2019. A novel sensors fault detection and self-correction method for HVAC systems using decentralized swarm intelligence algorithm. Int. J. Refrig 106, 54–65. https://doi.org/10.1016/j.ijrefrig.2019.06.007.

Yan, W., Guo, P., Gong, L., Li, Z., 2016. Nonlinear and robust statistical process monitoring based on variant autoencoders. Chemom. Intell. Lab. Syst. 158, 31–40. https://doi.org/10.1016/J.CHEMOLAB.2016.08.007.

Yi, T.-H., Huang, H.-B., Li, H.-N., 2017. Development of sensor validation methodologies for structural health monitoring: a comprehensive review. Measurement 109, 200–214. https://doi.org/10.1016/J.MEASUREMENT.2017.05.064.

Yin, X., Cheng, S., Yu, H., Pan, Y., Liu, Q., Huang, X., Gao, F., Jing, G., 2024. Probabilistic assessment of rockburst risk in TBM-excavated tunnels with multi-source data fusion. Tunn. Undergr. Sp. Technol. 152, 105915. https://doi.org/10.1016/j.tust.2024.105915.

Zhao, X., Liu, P., Mahmoudi, S., Garg, S., Kaddoum, G., Hassan, M.M., 2024. DDANF: deep denoising autoencoder normalizing flow for unsupervised multivariate time series anomaly detection. Alexandria Eng. J. 108, 436–444. https://doi.org/10.1016/j.aej.2024.07.013.