Article:
Koohmishi, M. and Connolly, D.P. (2025) Railway ballast fouling detection using thermal imaging: integration of LSTM and XGBoost. Transportation Geotechnics. 101889. ISSN: 2214-3912

https://doi.org/10.1016/j.trgeo.2025.101889

Railway Ballast Fouling Detection using Thermal Imaging: Integration of LSTM and XGBoost

Mehdi Koohmishi[1*], David P Connolly[2]

[1] Department of Civil Engineering, Faculty of Engineering, University of Bojnord, Bojnord, North Khorasan, Iran

[2] School of Civil Engineering, University of Leeds, Leeds, UK, United Kingdom

[*] Corresponding author: m.koohmishi@ub.ac.ir

Abstract

This paper presents an artificial intelligence (AI)-based approach to automate the structural health monitoring (SHM) of railway ballast through the fusion of long short-term memory and XGBoost (LSTM-XGB) to surface temperature data derived from infrared thermal images. In this context, machine learning models are trained using remotely acquired surface temperature data to classify fouling index based on thermal variations within ballast aggregates captured from thermograms. The long short-term memory (LSTM) component processes sequential time-series thermal data to predict preceding values, and the XGBoost (XGB) component classifies fouled ballast conditions based on identified patterns of surface temperature variations measured via infrared thermography (IRT). The results confirm the capability of the LSTM component to capture the time-series variations of a specimen's surface temperature in a shorter timeframe as well as the superior performance of XGBoost compared to a random forest (RF) approach, in classifying fouled ballast conditions. Therefore, the LSTM-XGB model demonstrates higher efficiency compared to the standalone XGBoost model, since the predictive nature of LSTM over time-series temperature data enables capturing shorter time

window for measuring ballast surface temperature and identifying patterns. Moreover, establishing a coarser classification of ballast fouling (categorized into three groups instead of five) significantly improves the model capability for accurate assessment of the ballast fouling conditions.

Keywords: Ballast fouling, Infrared thermography, LSTM, XGBoost, Random forest, Structural health monitoring

1 Introduction

1.1 Railway ballast structural health monitoring and infrared thermography

Railway ballast, typically composed of coarse crushed stone ranging from 20 to 63 mm in size, is distributed beneath and adjacent to the sleepers to ensure structural track stability and promote efficient drainage throughout the trackbed (Selig and Waters 1994). Among diverse non-destructive testing (NDT) methods employed for track structural health monitoring (SHM), using infrared thermography (IRT) has been proven as an effective technology for early damage detection within the superstructure and substructure of both ballasted and ballast-less railway tracks (see Fig. 1). With respect to the ballast layer, IRT stands out as an advanced technology capable of detecting fouled ballast layers in the field (Liang et al. 2023; Liang et al. 2024a; Koohmishi et al. 2024). IRT utilizes the thermal radiation characteristics of objects to map temperature distribution, offering advantages such as convenience, efficiency, large detection areas, and long-distance, non-contact operation (Yang et al. 2024). Notably, Aela et al. (2024) highlighted the effectiveness of vision-based techniques for monitoring railway superstructures, with IRT emerging as a suitable option for ballast layer inspection. For example, although ground penetrating radar (GPR) has the potential to depict the depth and thickness of subsurface irregularities (Guo et al. 2022), IRT is beneficial to obtain accurate horizontal measurements of voids (Schnebele et al. 2015).

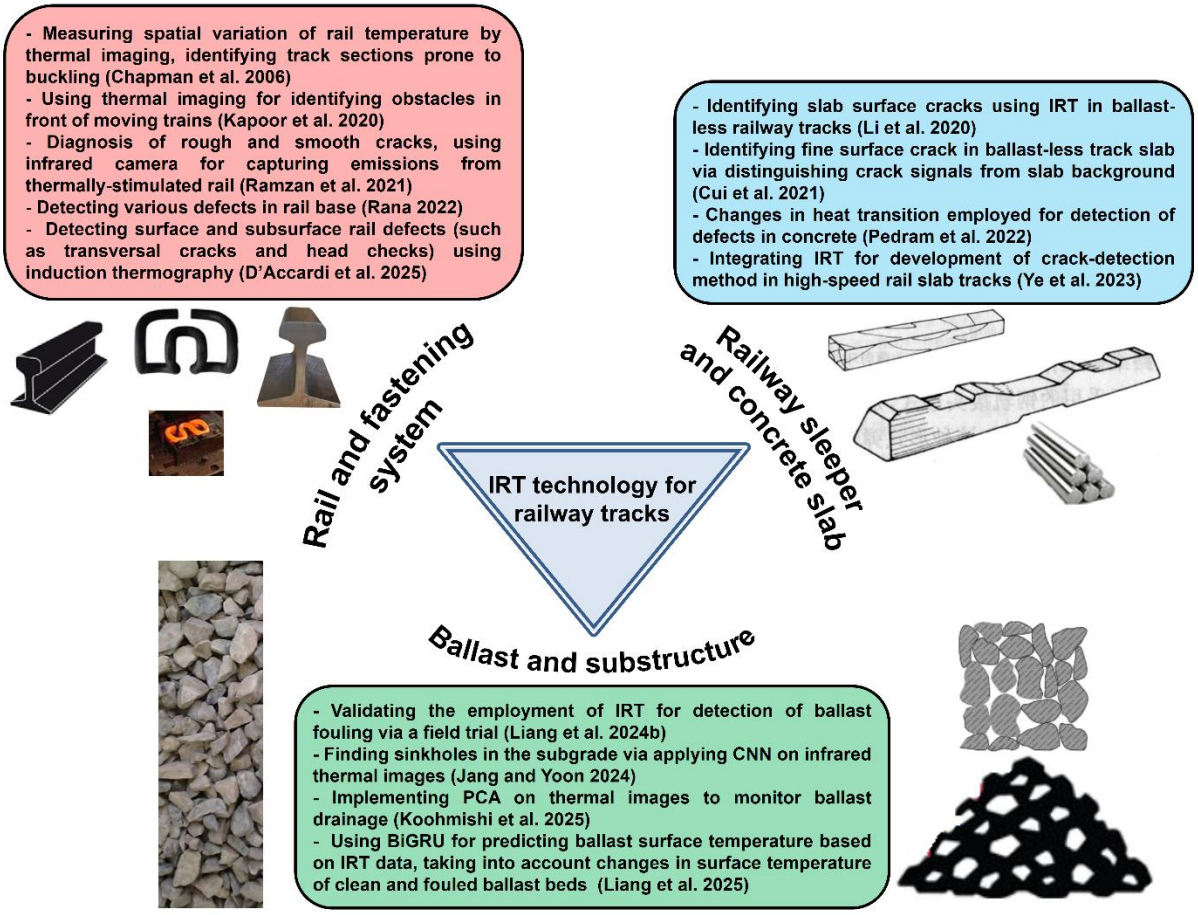


Fig. 1 Employment of IRT technology for SHM of distinct components of railway tracks (Chapman et al. 2006; Kapoor et al. 2020; Li et al. 2020; Cui et al. 2021; Ramzan et al. 2021; Pedram et al. 2022; Rana 2022; Ye et al. 2023; Jang and Yoon 2024; Liang et al. 2024b; Liang et al. 2025; D'Accardi et al. 2025; Koohmishi et al. 2025; Liang et al. 2025)

1.2 Processing infrared thermal images: A focus on AI integrating

With respect to performing statistical processing, the principle component analysis (PCA) underpinned by extensive data sources of IRT demonstrated ability to detect complex features exhibiting temperature variations (Koohmishi et al. 2025). Meanwhile, fusing artificial intelligence (AI) methods, such as machine learning (ML) and deep learning (DL) techniques, and IRT images have been conducted in earlier studies, as He et al. (2021) pointed out that ML and DL techniques contribute to automate thermal imaging-based visionary. In this regard, Xu and Hu (2021) employed gated recurrent model (GRU) on IRT images to identify defect depth, while PCA was conducted within the initial stage to reduce the dimension and correlation of

raw datasets. Additionally, Liang et al. (2025) used GRU model to predict surface temperature of ballast, aiding detection of ballast fouling.

1.3 Employment of ML and DL techniques for track health monitoring: A focus on LSTM, XGBoost, and RF models

Earlier studies (Guclu et al. 2010; Mishra et al. 2017) utilized statistical techniques, such as AutoRegressive-Moving-Average (ARMA) and Particle Filter-based technique, to assess the remaining useful life (RUL) of different components of the railway assets. Meanwhile, combining AI and statistical approaches has demonstrated rigorous potentials to enhance anticipatory aspects of track geotechnical properties (Hussaine et al. 2025; Rojas-Vivanco et al. 2025). More recently, ML and DL techniques have been proposed for characterized predictions. In this context, long short-term memory (LSTM) is characterized as a well-established method, particularly when measuring successive time-series data. Regarding rolling stock equipment, De Simone et al. (2023) pointed out superiority of LSTM with high-level forecasting accuracy, compared to other deep learning architectures to predict RUL. Taking into account the track superstructure and substructure, Chen et al. (2021) pointed out the appropriateness of LSTM to forecast the moisture dynamics in railway subgrades. Wang et al. (2023) integrated convolutional neural network (CNN) and LSTM on geometry data collected by inspection vehicle, while CNN was incorporated to capture the spatial dependence on adjacent segments. Again, prediction accuracy confirmed outperforming of characterized model compared to plain CNN, plain LSTM as well as multilayer perceptron. Recently, Deng et al. (2025) used time-series deformations derived from multi-modal data (space-air-ground monitoring) to integrate wavelet transform and LSTM for detecting points with potential sinking in karst areas along the railway tracks. Regarding classification models, random forest (RF) and extreme gradient boosting (XGBoost) are widely established procedures for categorizing the condition of railway tracks based on captured data. For instance, Wongkaew

et al. (2024) employed both models to characterize the buckling mode of rails using datasets derived from finite element analysis. Moreover, Alsahli and Alsulmi (2023) deployed RF and an artificial neural network on track geometry data to detect fouling levels ranging from clean to severe. Recently, Yürekli et al. (2025) utilized RF and XGBoost algorithms to enable real-time detection of railway-related disasters, including rockfalls, treefalls, and landslides. With respect to other ML/DL methods, Xiao et al. (2025a) used U-Net to identify surface fouling of railway ballast beds based on visual data.

2 Summary of research gap, contribution and methodology

Despite the use of thermograms for ballast health monitoring in earlier studies (Liang et al. 2023; 2024; Koohmishi et al. 2025), limited studies have explored the deployment of ML/DL techniques on IRT as a non-invasive method for railway track monitoring. These techniques aim to accelerate the detection of subsurface defects (particularly ballast fouling) from surface thermal images, as well as to support the ongoing classification of fouled ballast. Therefore, the present study seeks to classify fouled ballast conditions using AI-enhanced passive IRT and to reduce the time required for on-board SHM of ballasted tracks. To achieve this, thermal images captured horizontally from the ballast specimen surface are utilized to analyze temperature distribution, while also providing time-series datasets for the development of ML/DL techniques through the fusion of LSTM and XGBoost or RF models, aiming to predict long-term variations and identify patterns. Moreover, statistical analysis is conducted on surface temperature variations observed in clean and fouled ballast specimens during cooling cycles.

3 Materials and test setup

3.1 Railway ballast and fouling materials

115 Table 1.a details the characteristics of virgin ballast aggregates sourced from a raw material
116 site. Moreover, Table 1.b outlines the geotechnical attributes of the clayey soil employed as a
117 fouling agent. In the current study, clay was incorporated into the ballast layer as an extremely
118 derived contaminant, classified as a fouling material known to impair the fluid flow capacity
119 of the granular matrix (Ramadan et al. 2025). Fig. 2 illustrates a prepared specimen of the clean
120 ballast layer prior to contamination.

121 **Table 1** Physical and mechanical properties of virgin ballast and fouling materials

a Ballast aggregate			**b** Fouling material	
Bulk dry Specific gravity		2.72	Specific gravity	2.58
Color		Gray	Plastic limit (%)	17
Water absorption (%)		0.56	Liquid limit (%)	39
AREMA No. 25	d_{max} (mm)	62.5	d_{max} (mm)	0.42
	d_{60} (mm)	41	d_{60} (mm)	0.047
	d_{10} (mm)	15	d_{50} (mm)	0.018

122

123

Fig. 2 Prepared clean ballast specimen for measuring surface and subsurface temperatures using IRT and thermometers

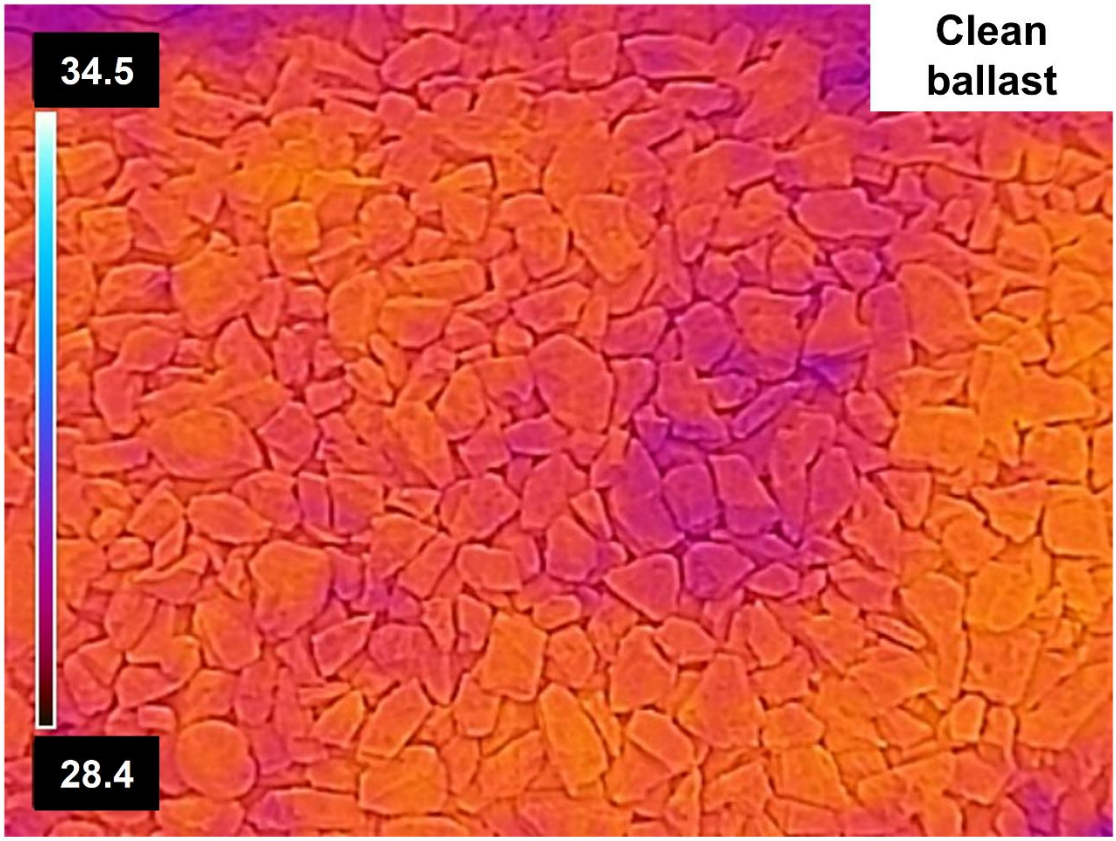
3.2 Thermal camera and thermometers

An infrared thermal imaging device, depicted in Fig. 3.a.1, operating within the $7.5 - 13$ μm spectral range, was employed to acquire thermal profiles of ballast specimens in both clean and clay-fouled configurations. The device offers a resolution of 96×128 pixels and a temperature detection interval spanning -20 ºC to 150 ºC, with measurement accuracy of ± 0.1 ºC. Regarding one-meter distance between camera and ballast specimen's surface and horizontal field of view of 53.6 º, the spatial image resolution of the thermal image is 7.89×7.89 mm/pixel. The emissivity setting was calibrated using the manufacturer's recommended value (FLIR 2014) and finally the value of 0.95 was established for emissivity regarding ballast surface temperature measurement using contact thermometer and captured temperature using IRT. To validate IRT measurements, thermometers were deployed at designated control points (see Figs. 3.a.2). Fig. 3.b shows a representative thermal image of prepared ballast specimen, enabling detailed analysis of temperature across the granular matrix. Fig. 3.c provides a schematic overview of the specimen layout, indicating the precise placement of thermometers used for calibration and verification.

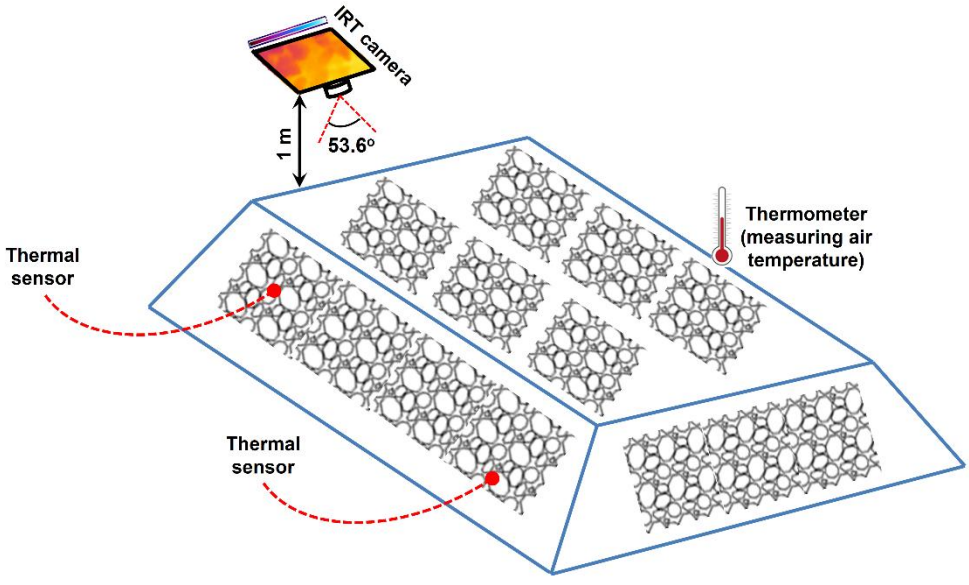
a.1 Thermal camera (FLIR)

a.2 Thermometer and hygrometer

a Instruments used for conducting the experimental program



b Thermal image of clean ballast specimen



c Schematic illustration of ballast specimen and placement of thermal sensors

Fig. 3 FLIR camera used for capturing thermal images of ballast specimens

With respect to relying on sunlight as a heat source, commonly specified as the passive approach, the datasets used to train LSTM-XGB and LSTM-RF models were all collected

during the summer, when optimal thermal excitation and pronounced thermal contrast can be effectively captured via IRT technology (Pedram et al. 2022; Pozzer et al. 2024). Therefore, despite observing more noticeable variations in surface temperature of water-filled voids (Liang et al. 2023; Jang et al. 2023), the effects of rainfall and water presence within granular particles have not been accounted for.

3.3 Quantification of external-sourced ballast fouling agents

To quantify the degree of clay intrusion, characterizing externally sourced fouling embedded in the ballast, the fouling index is computed as detailed below (Selig and waters 1995):

$$FI = P_{4.75} + P_{0.075} \quad (1)$$

FI = Fouling index (%)

$P_{4.75}$= Material passing through the 4.75 mm sieve (%)

$P_{0.075}$= Material passing through the 0.075 mm sieve (%)

Regarding the classification presented in Table 2 with respect to fouling levels in the ballast layer, varying FI values (ranging from 5% to 50%), together with the clean ballast samples, were fabricated to simulate diverse conditions, as instances are depicted in Fig. 4. In addition to the detailed categorization in Table 2, a coarser classification scheme for ballast fouling was also established, comprising three ranges: (I) $FI < 1\%$; (II) $1\% \leq FI < 20\%$; and (III) $FI \geq 20\%$.

Table 2 Classification of ballast fouling based on the FI and percentage of fouling (Selig and Waters 1995; Indraratna et al. 2011)

Category	Designation	FI (%)	Percentage of fouling[1] (%)
Clean	I	< 1	< 2
Moderately clean	II	1 to < 10	2 to < 9.5

Moderately fouled	III	10 to < 20	9.5 to < 17.5
Fouled	IV	20 to < 40	17.5 to < 34
Highly fouled	V	≥ 40	≥ 34

[1] Ratio of the dry weight of material passing the 9.5 mm sieve to the total dry weight of the sample

a FI $= 15\%$

b FI $= 30\%$

c FI $= 45\%$

Fig. 4 Clay-fouled ballast particles considering different FI values

4 Development of AI-based model for analysis of thermal images

4.1 Statistical analysis of thermal images

To analyze thermal images captured from the ballast surface, the average and standard deviation of pixel temperatures in the thermogram are measured as follows:

$$\mu_T = \frac{1}{M.N}\sum_{i=1}^{M}\sum_{j=1}^{N} T_{i,j} \quad (2)$$

$$Std_T = \sqrt{\frac{1}{M.N}\sum_{i=1}^{M}\sum_{j=1}^{N}(T_{i,j} - \mu_T)^2} \quad (3)$$

$T_{i,j}$= Temperature for pixel (i,j)

μ_T= Temperature averaged over all thermogram pixels

Std_T= Standard deviation of pixel temperatures across the thermogram

M = Number of rows in thermal image

N = Number of columns in thermal image

4.2 General explanation of LSTM-RF and LSTM-XGB models

The LSTM-RF and LSTM-XGB models developed in this study comprise a deep learning component, called long short-term memory (LSTM), to predict the surface temperature of ballast specimens based on time series data measured via IRT, as well as a machine learning component, RF or XGBoost, to classify ballast fouling based on the statistical features of the measured and predicted surface temperatures and air temperature. During diurnal temperature fluctuations in the cooling down timeframe (3:30 PM to 7:30 PM) over a 60-day period, a total of 2880 sequential thermal images captured from the surface of ballast specimens with characterized FI levels were integrated into the AI-based model. Fig. 5 shows a schematic layout of the established LSTM-RF and LSTM-XGB models, while further details are provided in the following subsections.

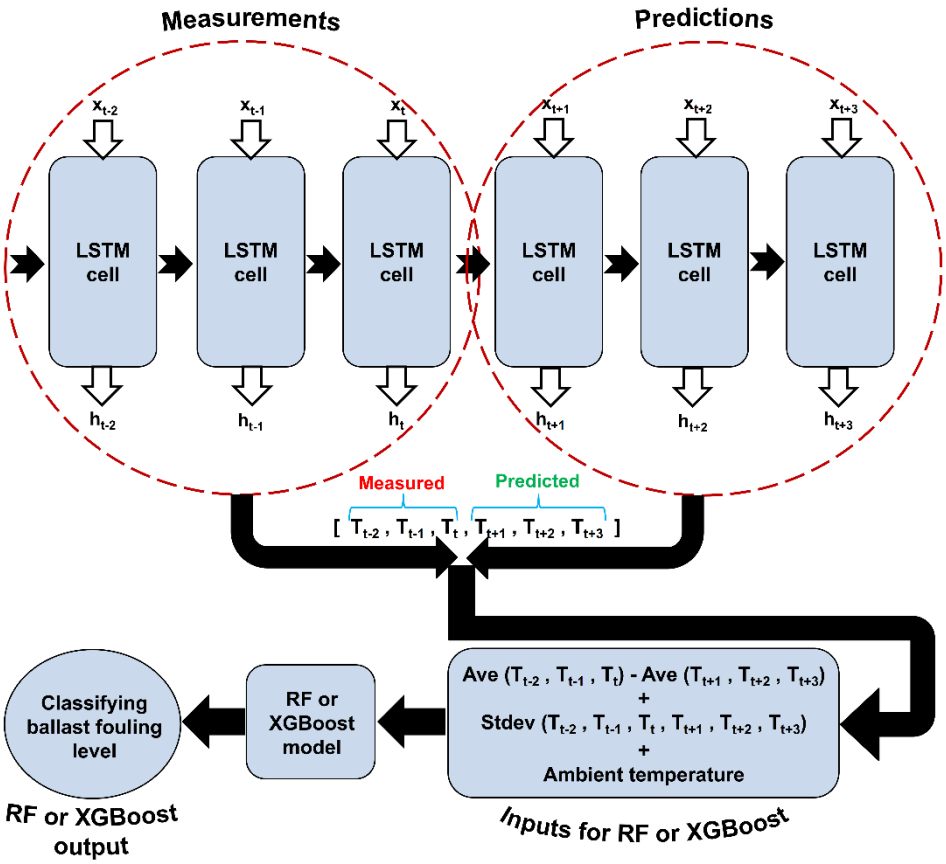


Fig. 5 Two-stage pipelines of LSTM-RF/LSTM-XGB models: Decoupling temporal temperature component from fouling classification constituent via fusion of LSTM and RF (or XGBoost)

4.3 LSTM model

LSTM is characterized as a promoted version of recurrent neural network (RNN), aiming to address the issue of gradient explosion and gradient vanishing (Hochreiter and Schmidhuber 1997). The dominant gates of LSTM structure are defined as follows (as illustrated in Fig. 6):

$$I_t=\sigma(W_{xi}x_t+W_{hi}h_{t-1}+b_i) \quad (4)$$

$$F_t=\sigma(W_{xf}x_t+W_{hf}h_{t-1}+b_f) \quad (5)$$

$$O_t=\sigma(W_{xo}x_t+W_{ho}h_{t-1}+b_o) \quad (6)$$

$$\hat{C}_t=\tanh(W_{xc}x_t+W_{hc}h_{t-1}+b_c) \quad (7)$$

$$C_t=F_t \otimes C_{t-1} + I_t \otimes \hat{C}_t \quad (8)$$

$$h_t=O_t \otimes \tanh(C_t) \quad (9)$$

206 I_t= Input gate

207 F_t= Forget gate

208 O_t= Output gate

209 C_t= Gate control unit (cell state) at time step t

210 h_{t-1}= Hidden state from previous time step (Output value at t-1)

211 h_t= Hidden state (output) value at time step t

212 x_t= Input value at time step t

213 C_{t-1}= Gate control unit (cell state) from previous time step (at t-1)

214 $\hat{C}_t$= Candidate cell state at time step t

215 b_i= Bias term for input gate

216 b_f= Bias term for forget gate

217 b_o= Bias term for output gate

218 b_c= Bias term for cell state

219 W_{xi}= Weight matrix of input gate for input x_t

220 W_{xf}= Weight matrix of forget gate for input x_t

221 W_{xo}= Weight matrix of output gate for input x_t

222 W_{xc}= Weight matrix of cell state update for input x_t

223 W_{hi}= Weight matrix of input gate for previous hidden state h_{t-1}

224 W_{hf}= Weight matrix of forget gate for previous hidden state h_{t-1}

225 W_{ho}= Weight matrix of output gate for previous hidden state h_{t-1}

226 W_{hc}= Weight matrix of cell state update for previous hidden state h_{t-1}

227 $\sigma(i)$= Sigmoid activation function

228 $\tanh(i)$= Hyperbolic tangent activation function

229 Forget gate captures no further useful information that should be discarded from the cell

230 state. Input gate characterizes new information that should be stored in the cell state

representing filtration and integration of new information. Finally, output gate decides about part of the cell state to become as hidden state for the next step's context (Hochreiter and Schmidhuber 1997; Gers and Schmidhuber 2001). Fig. 7 shows a sequence of time-series thermal images (capturing six timeframes) used for the deployment of LSTM model.

Regarding the major hyperparameters of established model, the hidden size (units), number of layers, dropout value, and learning rate, are tuned to find out the most appropriate structure of LSTM model. With respect to ensuring the LSTM model could learn from a series of previous time steps, the dataset is then divided into training and testing sets, 80% of the dataset is set aside for training, while the remaining 20% is preserved for testing (Xiao et al. 2019).

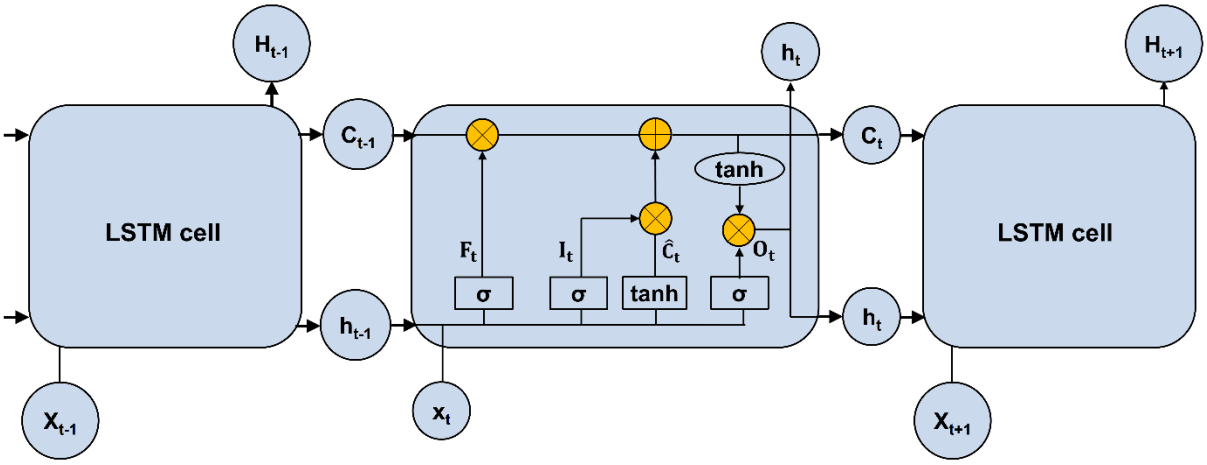


Fig. 6 LSTM applied on time-series datasets of ballast surface temperature measurements derived from IRT

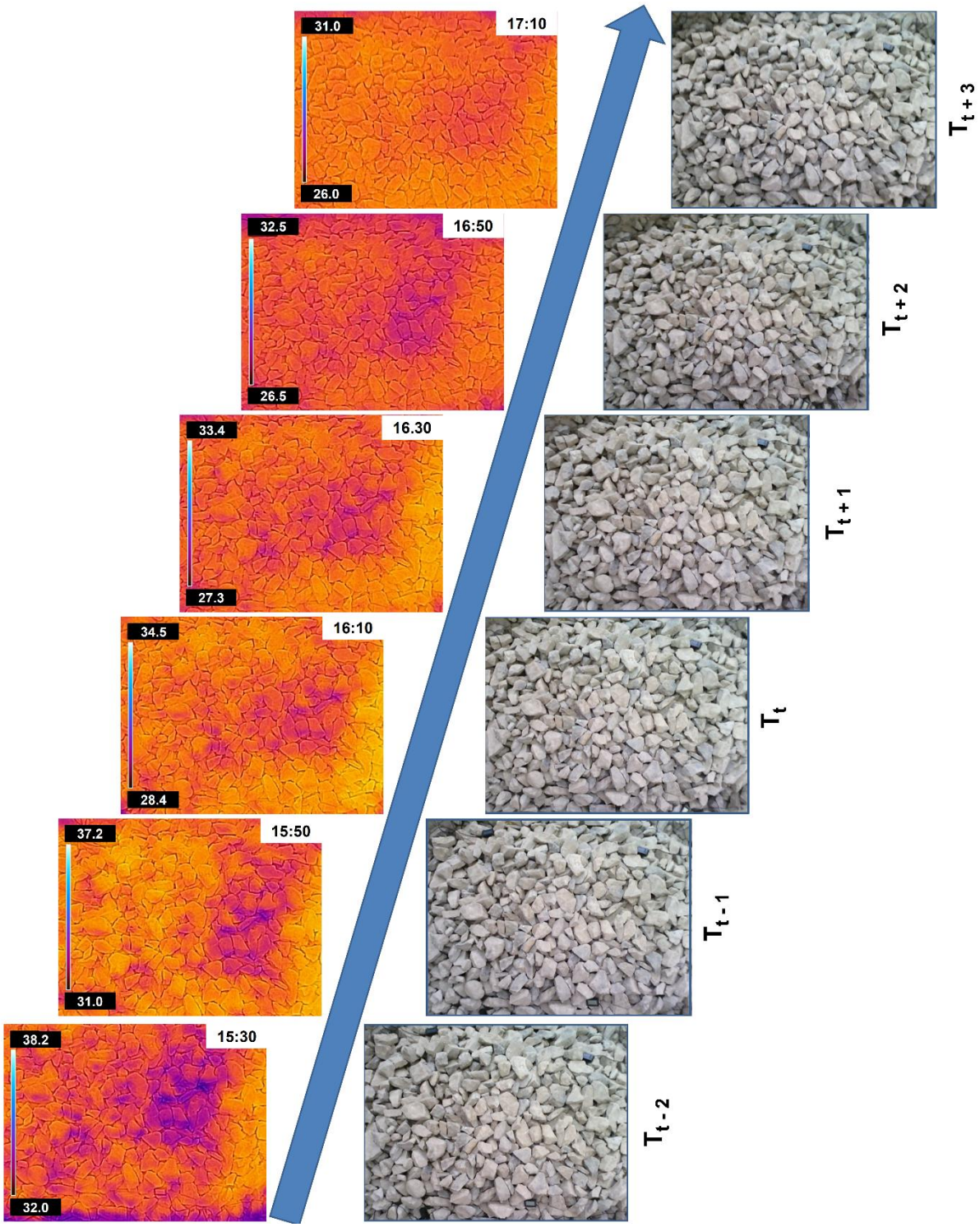


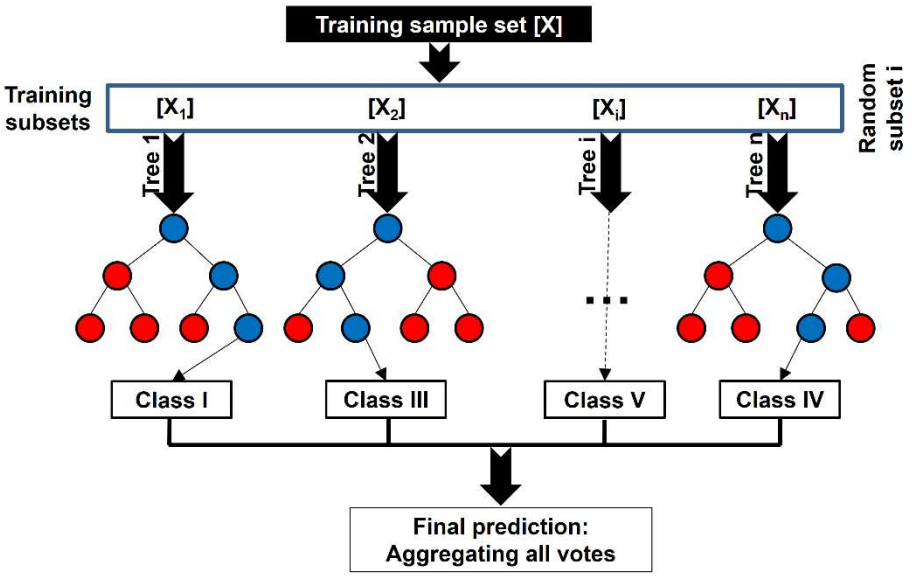
Fig. 7 Temporal thermal images employed for learning LSTM component of LSTM-RF and LSTM-XGB models

4.4 RF and XGBoost models

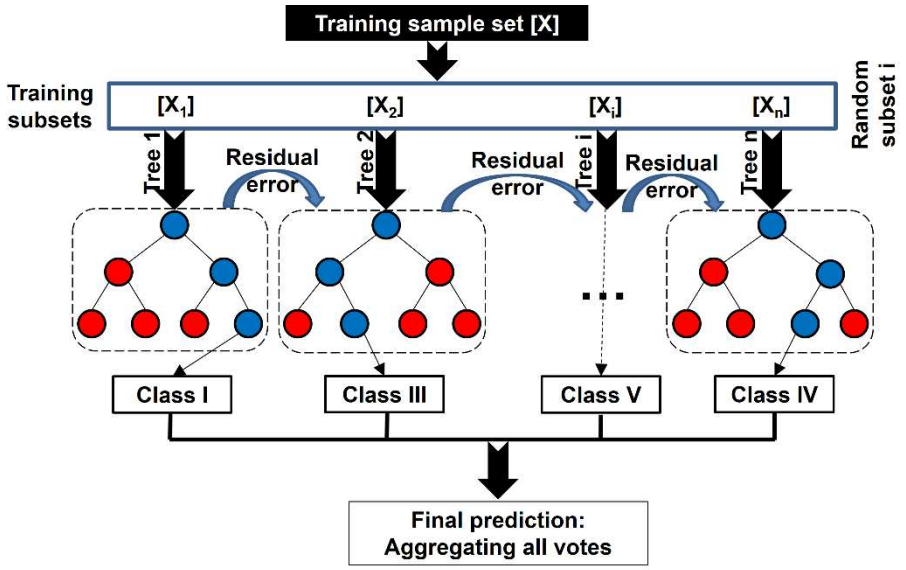
- ***Random forest (RF) model***. RF classification approach (Breiman 2001) builds upon the creation of an ensemble of decision trees, each trained on a stochastic fraction of the overall characterized dataset. The class label is predicted through majority voting, based on the aggregated votes from all individual trees. Fig. 8.a shows an illustrative layout of the RF classifier established in the present study to categorize ballast fouling according to the labels provided in Table 2. The major hyperparamters—including number of trees in the forest (n_estimators), maximum depth of each tree (max_depth), as well as maximum number of features established (max_features)—are tuned to find out the most appropriate structure of RF model.

- ***Extreme gradient boosting (XGBoost) model***. The XGBoost classifier, pioneered by Chen and Guestrin (2016), constitutes a robust ensemble ML approach, wherein an ensemble of decision trees in a consecutive manner is built, each trained to rectify the misclassifications made by the preceding ones. The key parameters in XGB classifier include the number of boosting rounds (n_estimators), maximum depth of each tree (max_depth), learning rate representing shrinkage weight applied to each tree (eta), and fraction of training data used for individual tree (subsample). Fig. 8.b illustrates the structure of XGBoost established for classification of ballast fouling.

To ensure consistency within time-series temperature data, average and standard deviation of surface temperature acquired by thermal images are deployed as dominant inputs for RF and XGBoost classifier. Meanwhile, the air temperature is used as another input for RF and XGBoost models, as earlier study (AlDousari et al. 2024) pointed out the presence of a strong positive relationship between solar irradiance and surface temperature. Moreover, aiming to learn the model, the collected samples are divided into the training and testing with proportions of 70% and 30%, respectively (Farhadi and Najafzadeh 2021).



a Structures of RF constituent of LSTM-RF model



b Structures of XGBoost constituent of LSTM-XGB model

Fig. 8 Structures of RF and XGBoost models applied on statistical features of time-series ballast surface temperature to classify ballast fouling

4.4 Evaluation indicators

To evaluate the appropriateness of established models, including LSTM predictions as well as RF and XGBoost classifications, diverse employed indices are determined as follows. Overall, the LSTM performance is evaluated by computing RMSE, MAE, EV, and R^2 across multiple-step outputs due to capturing multiple successive time steps over predictions (Wakjira et al.

2022). Meanwhile, the RF performance is assessed via determination of accuracy, precision, recall, and F1 Score.

- Root mean square error (RMSE). Indicting the sum of squares of differences between the predicted and real values:

$$RMSE=\sqrt{\frac{1}{n.m}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{i,j}-\hat{y}_{i,j})^2} \quad (10)$$

- Mean absolute error (MAE). Indicting the average of sum of absolute error values (difference between the predicted and real values):

$$MAE=\frac{1}{n.m}\sum_{i=1}^{n}\sum_{i=1}^{m}\left|y_{i,j}-\hat{y}_{i,j}\right| \quad (11)$$

- Explained variance (EV). Representing the level of capturing variations in the real values by predictions:

$$EV=\left(1-\frac{\sum_{i=1}^{m}Var\left|Y_j-\hat{Y}_j\right|}{\sum_{i=1}^{m}Var(Y_j)}\right)\times 100 \quad (12)$$

- Coefficient of determination (R^2). Reflecting the degree of explaining dependent variable based on the independent variable:

$$R^2=1-\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{i,j}-\hat{y}_{i,j})^2}{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{i,j}-\overline{Y}_j)^2} \quad (13)$$

To determine EV and R^2 as a single metric for general performance of model, a flattened vector of all actual and predicted values is employed. The variables utilized in Eqs. (10-13) are defined as follows:

$y_{i,j}$= Actual value for sample i, step j

$\hat{y}_{i,j}$= Predicted value for sample i, step j

Y_j= Vector of actual values for output j

$\hat{Y}_j$= Vector of predicted values for output j

$\overline{Y}_j$= Average across all actual values of output j

$Var(Y_j)$= Variance across all actual values of output j

n= Total number of samples

m= Number of outputs

- *Accuracy.* Capture proportions of correctly classified predictions out of the total samples (overall correctness) as follows:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

- *Precision and recall.* Precision determines the proportion of predicted positives that are correct, while recall measures the proportion of actual positives that are correctly identified:

$$\text{Precision}=\frac{TP}{TP+FP} \quad (15)$$

$$\text{Recall}=\frac{TP}{TP+FN} \quad (16)$$

- *F1 Score.* A metric that harmonizes both precision and recall, particularly when both false positives and false negatives are important, defined by the following formulas:

$$\text{F1 Score}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (17)$$

Where the definitions of variables used in Eqs. (14-17) are provided below:

TP= True positive

TN= True negative

FP= False positive

FN= False negative

5 Results and discussions

5.1 Effects of fouling level on ballast surface temperature captured via IRT

- *Mean surface temperature.* To investigate the effect of fouling level on the statistical characteristics of surface temperature in clay-contaminated ballast specimens, Fig. 9 presents

average values across a range of ambient temperatures. With respect to clean and fouled ballast specimens, the thermal inertia of the ballast increases with higher fouling levels, which slows down the cooling cycles and results in elevated surface temperature values for fouled specimens (FI levels of 15%, 30%, and 45%). This delayed thermal response is primarily attributed to a slower reaction to ambient temperature changes, leading to distinct thermal signatures in fouled zones compared to clean ballast. Regarding the detection of voids in the road pavement subgrade, Jang et al. (2023) reported a similar trend for the original subgrade, noting a stable thermal profile in extremely compacted layers and observable temperature fluctuations in void-filled subgrade layers.

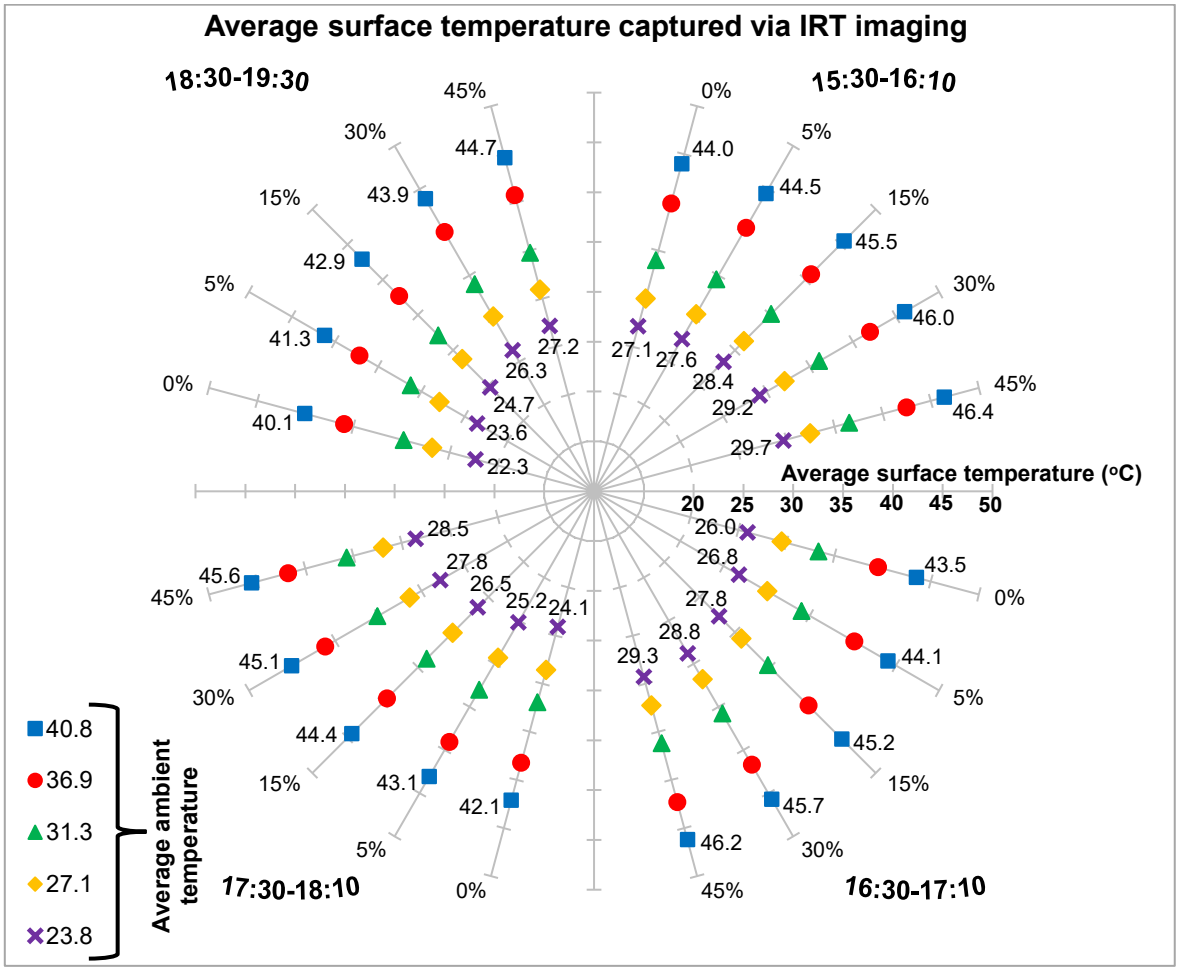
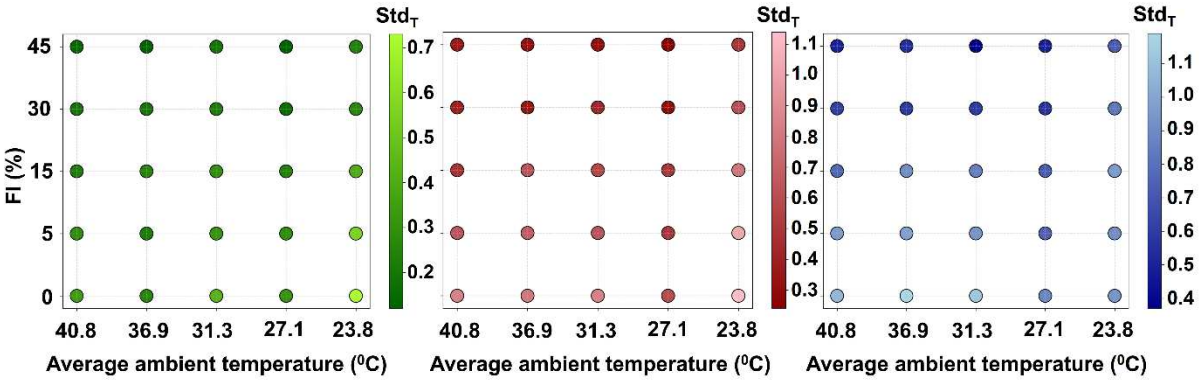


Fig. 9 Average surface temperature of fouled ballast specimens captured using thermal images

- ***Standard deviation of surface temperature.*** Fig. 10 illustrates the standard deviation values of surface temperature (Std_T), captured via thermal imaging across distinct FI levels, confirming a reduction in Std_T with increasing fouling severity. This trend reflects the greater thermal inertia of clay compared to large-sized ballast particles; consequently, sharper temperature transitions are observed in clean ballast than in clay-fouled ballast specimens. The low thermal diffusivity of clay materials causes clay-contaminated ballast to retain thermal inertia, resulting in prolonged temporal correlations in temperature and reduced variability over time. As Anglade et al. (2022) noted, clay-rich samples exhibit lower thermal conductivity and higher heat capacity than sand-rich ones, both contributing to their greater thermal inertia.



a 15:30 - 17:10 **b** 16:30 - 18:10 **c** 17:30 - 19:10

Fig. 10 Standard deviation of surface temperature in fouled ballast specimens obtained from thermal images

- ***GLCM of thermal images.*** To provide a detailed analysis of temperature variations within captured thermal images from ballast's surface, gray level co-occurrence matrix (GLCM) is computed for each specific patch across thermogram to measure local texture features like homogeneity and contrast. Fig. 11 illustrates the mean values of these characteristics across the image of ballast's surface across diverse ranges of fouling level. AS illustrated, increase in FI leads to higher homogeneity and lower contrast, associated with more uniform texture and

lower local intensity variations. Therefore, the established IRT clusters fouling categories potentially through discriminating fouling based on GLCM features.

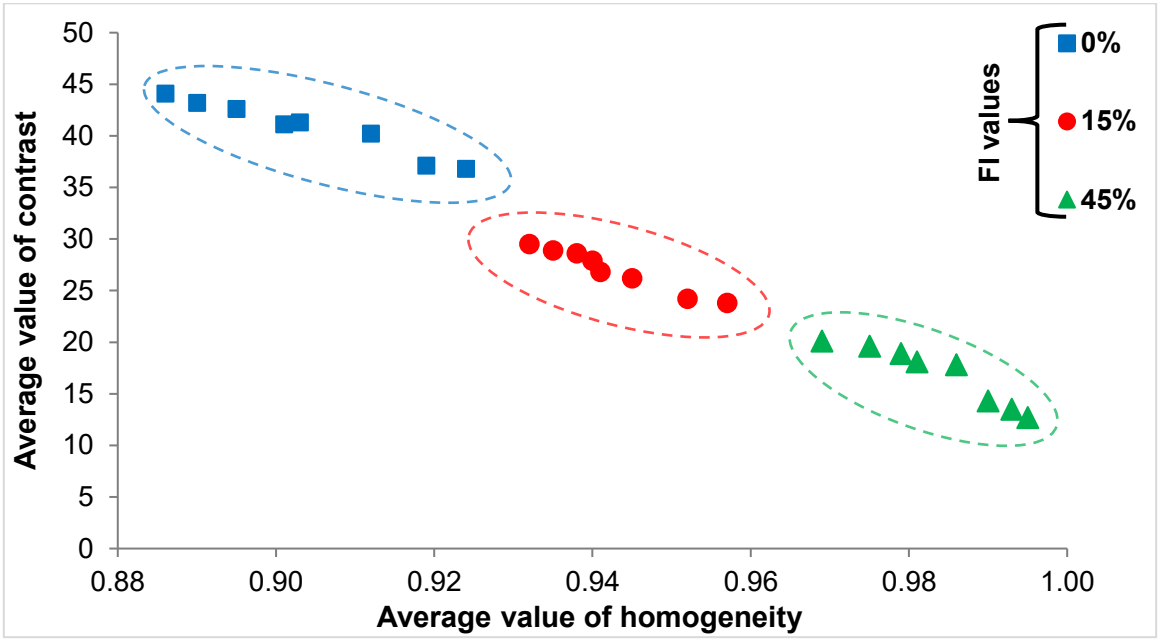


Fig. 11 Homogeneity and contrast values derived from GLCM of thermal images of ballast specimens with diverse FI values

5.2 Results of AI-based models for analysis of thermal images

5.2.1 Performance of characterized AI-based models

- ***Evaluation indicators.*** Regarding established indicators for assessing the performance of LSTM, RF and XGBoost models, Table 3 presents the values of indices. Generally, the LSTM component of the developed model demonstrates high level of accuracy to predict surface temperature when three successive thermal images are integrated for predicting three subsequent temperatures (according to Fig. 5). Likewise, the classification component of the developed model signifies strong discriminative power between diverse levels of ballast fouling when establishing three specific classes rather than five ones. In this context, Table 4 presents the dominant configurations of LSTM and XGBoost models, as hyperparameter tuning was employed to derive their key parameters. The multi-output LSTM predicts three

future temperatures (outputs) based on three past temperatures (inputs). Moreover, Fig. 12 shows the confusion matrix based on the RF and XGBoost classifiers, demonstrating the outperformance of XGBoost for accurate classification of fouled ballast specimens, as dominant data are aligned on the diagonal of characterized matrix. For instance, XGBoost predicts correctly 238 samples from all 300 testing dataset provided for class II, while correct classification is 220 for RF classifier. From this tabular matrix, the values of accuracy, precision, recall, and F1 Score, can be calculated as presented in Table 3. As illustrated, three classes are established, including $FI < 1\%$, $1\% \leq FI < 20\%$, and $FI \geq 20\%$, because the differences in surface temperatures captured via IRT are not meaningful when five categories (according to Table 2) are defined.

Table 3 Performance of LSTM for sequential time-series surface thermal data as well as RF and XGBoost for classification of ballast fouling based on defined metrics

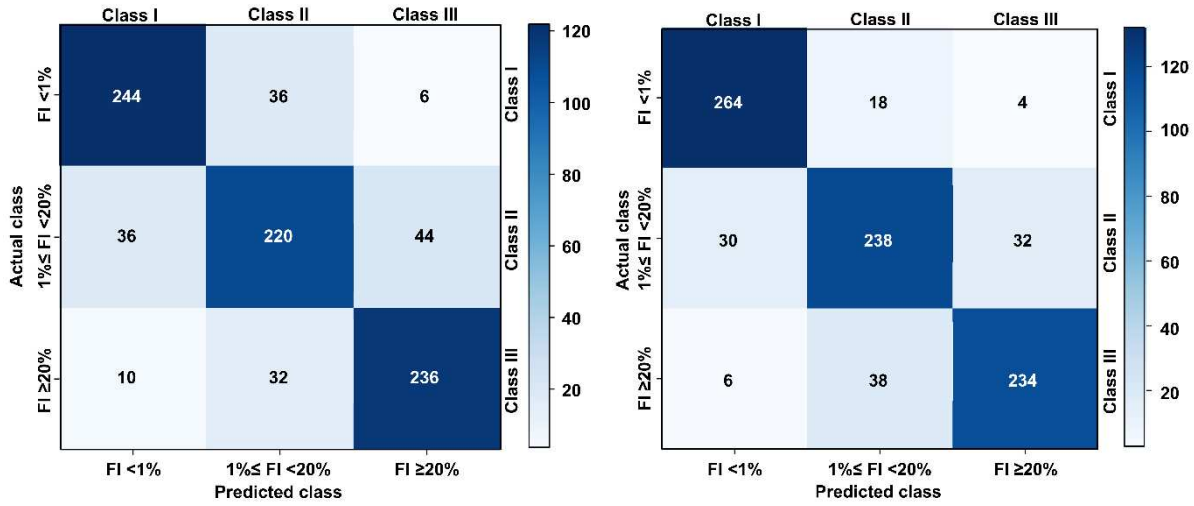
LSTM			RF classification			XGBoost classification		
	3 inputs and 3 outputs	5 inputs and 1 outputs		3 classes	5 classes		3 classes	5 classes
RMSE	0.409 / 0.495	0.128 / 0.145	Accuracy	0.931 / 0.810	0.848 / 0.744	Accuracy	0.975 / 0.852	0.868 / 0.778
MAE	0.330 / 0.406	0.096 / 0.099	Precision	0.930 / 0.809	0.849 / 0.743	Precision	0.975 / 0.851	0.870 / 0.775
EV	0.994 / 0.995	0.999 / 0.999	Recall	0.931 / 0.810	0.848 / 0.744	Recall	0.975 / 0.852	0.868 / 0.778
R^2	0.993 / 0.989	0.999 / 0.999	F1 Score	0.930 / 0.810	0.848 / 0.742	F1 Score	0.975 / 0.851	0.868 / 0.775

* Training / Testing

Table 4 Configuration of structures established for LSTM and XGBoost models

LSTM prediction	XGBoost classification

Feature	Value	Feature	Value
1st layer (Dropout)	32 (0.1)	n_estimators	250
2nd layer (Dropout)	16 (0.1)	learning_rate	0.05
learning_rate	0.01	reg_alpha, reg_lambda	0.05 , 2
bacth_size	20	max_depth	6
epochs	50	subsample	0.8



a RF classifier **b** XGB classifier

Fig. 12 Confusion matrix exhibiting the RF and XGBoost performance for classifying ballast based on fouling levels within three defined classes [testing dataset]

- Effect of timeframe established for time-series temperature measurement. With respect to the developed LSTM component, a heatmap is provided in Fig. 13, demonstrating the influence of time window as input and successive time steps for prediction based on MAE. As illustrated, the longer time window as input and a shorter predicted timeframe lead to lower values of MAE. Finally, a time window of three and predicted time steps of three were established to develop LSTM model (with an MAE value of 0.406). Regarding the 20-minute time interval between successive thermal image captures, using only three measurements (rather than more ones) can significantly reduce the time required for SHM of the ballast layer, as Loche et al.

(2022) found out that a 10-minute cooling rate was less time-consuming compared to integrating the entire cooling period for assessing the physical and mechanical properties of soils via IRT technology. Additionally, the autocorrelation values presented in Fig. 14 confirms the strong correlation at low lags, properly capturing short-term dependencies of surface temperature.

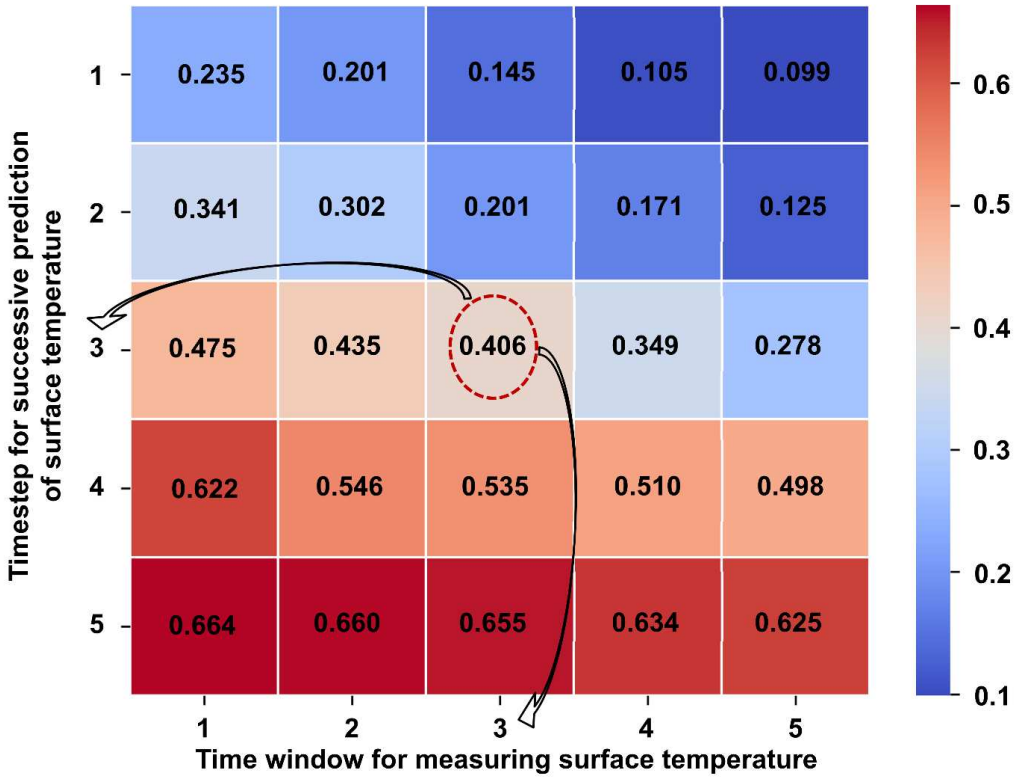
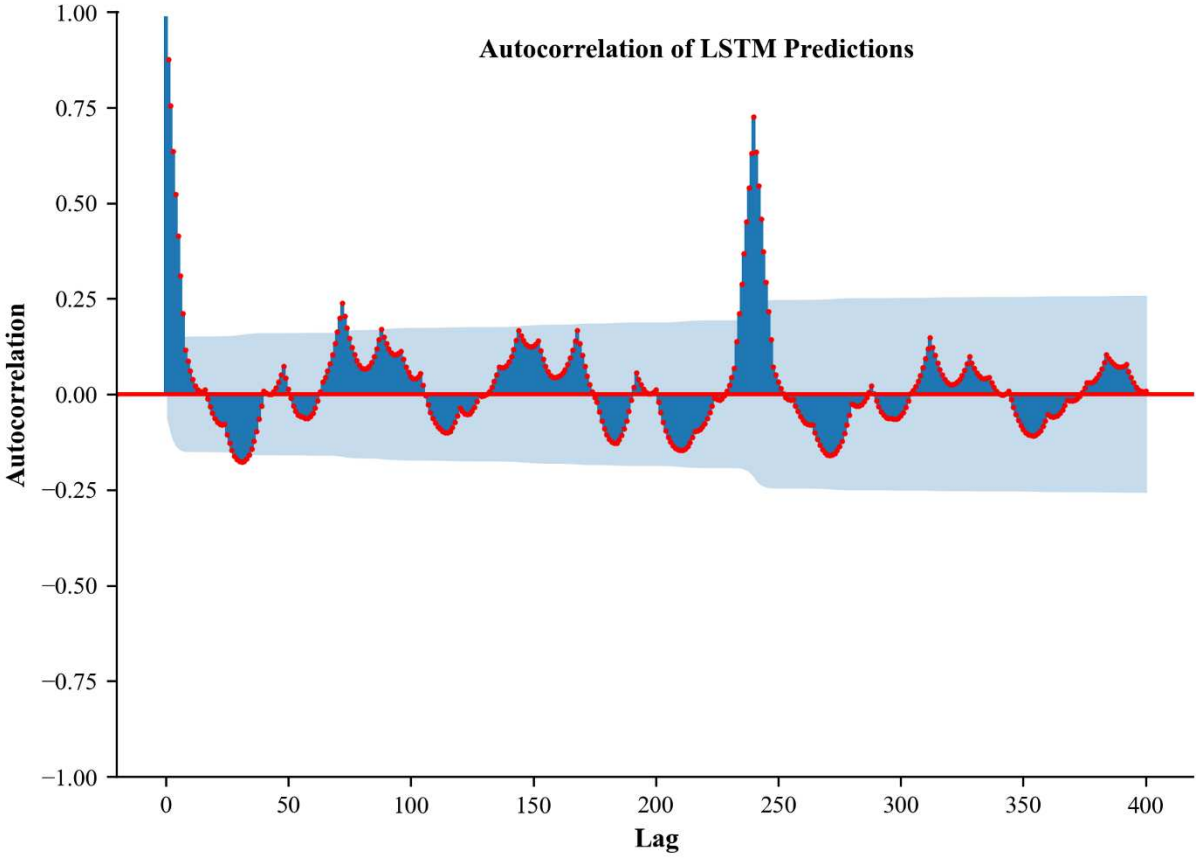
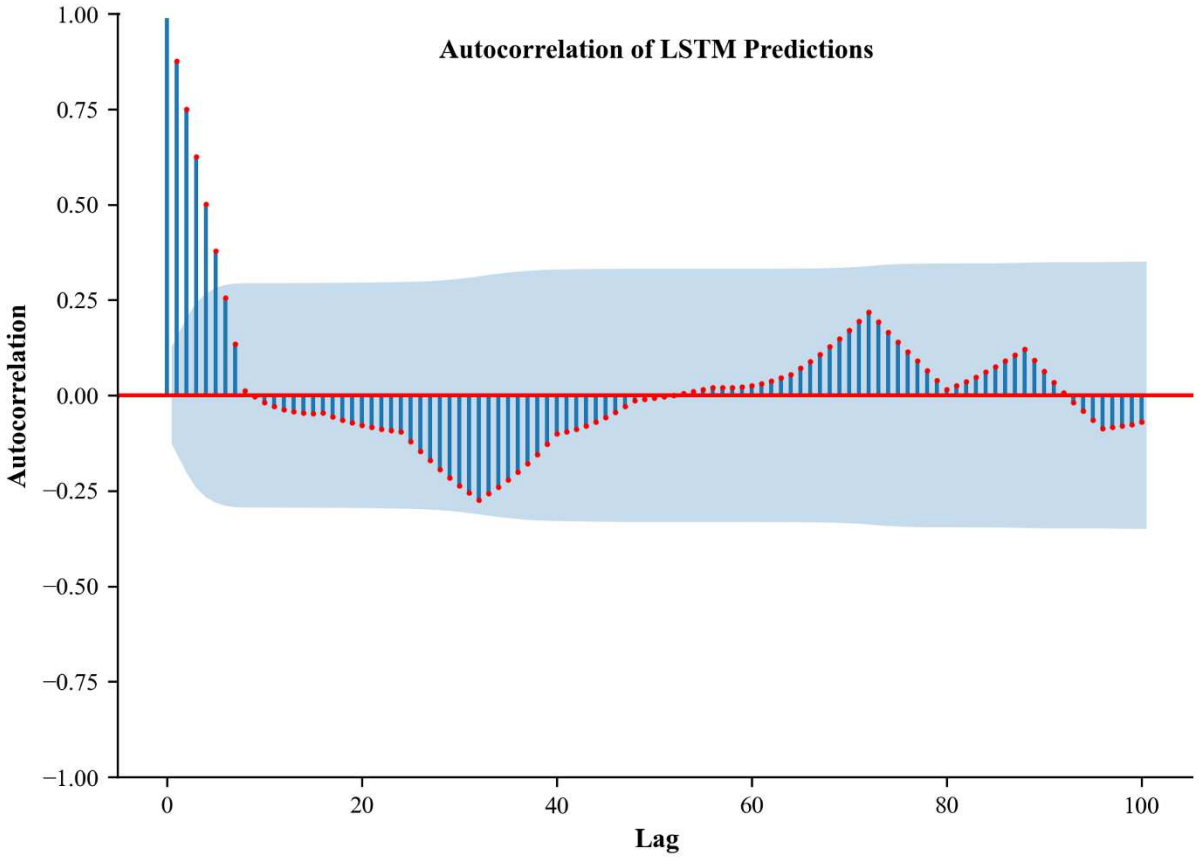


Fig. 13 Heatmap representing effects of various Tin/Tout choices established for LSTM component of LSTM-XGB model on MAE [testing dataset]



418

419 **a** Training dataset

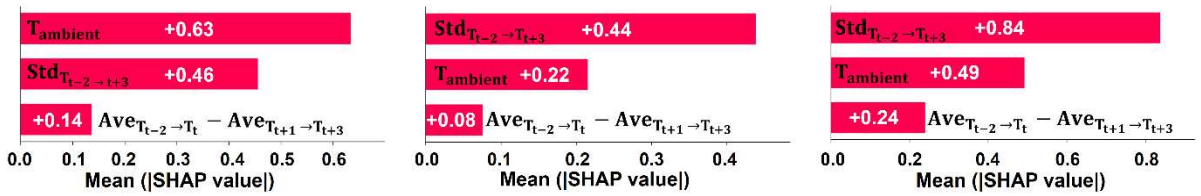


420

b Testing dataset

Fig. 14 Autocorrelation of LSTM predictions

- Feature importance of characterized input parameters for classification. To assess which features most influence the model's classification, Fig. 15 presents the mean absolute SHAP value for each feature across all categorized fouling levels. For clean ballast, ambient temperature is the dominant feature, whereas the standard deviation of ballast surface temperature over six successive readings becomes the dominant factor for moderately and heavily fouled specimens. Clean ballast permits efficient heat transfer, resulting in a close relationship with air temperature, while fouling agents within the granular particles dampen variability in surface temperature, leading to a reduced spread across successive readings.



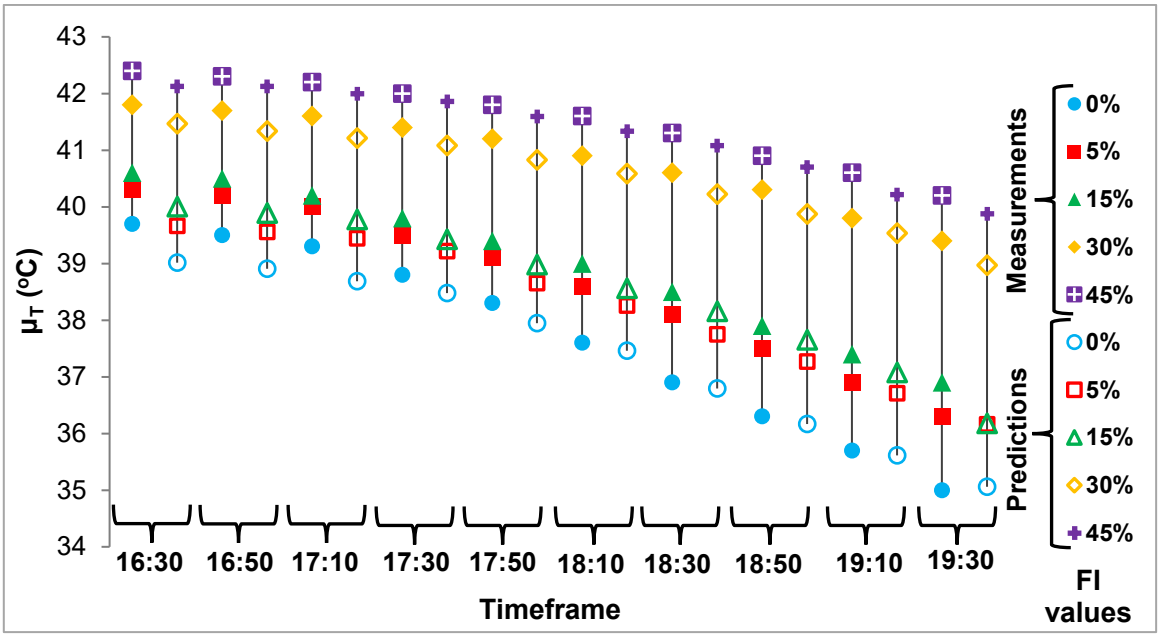
a Class I (FI $<$ 1%) **b** Class II (1% $\leq$ FI $<$ 20%) **c** Class III (FI $\geq$ 20%)

Fig. 15 Feature importance based on developed XGBoost model

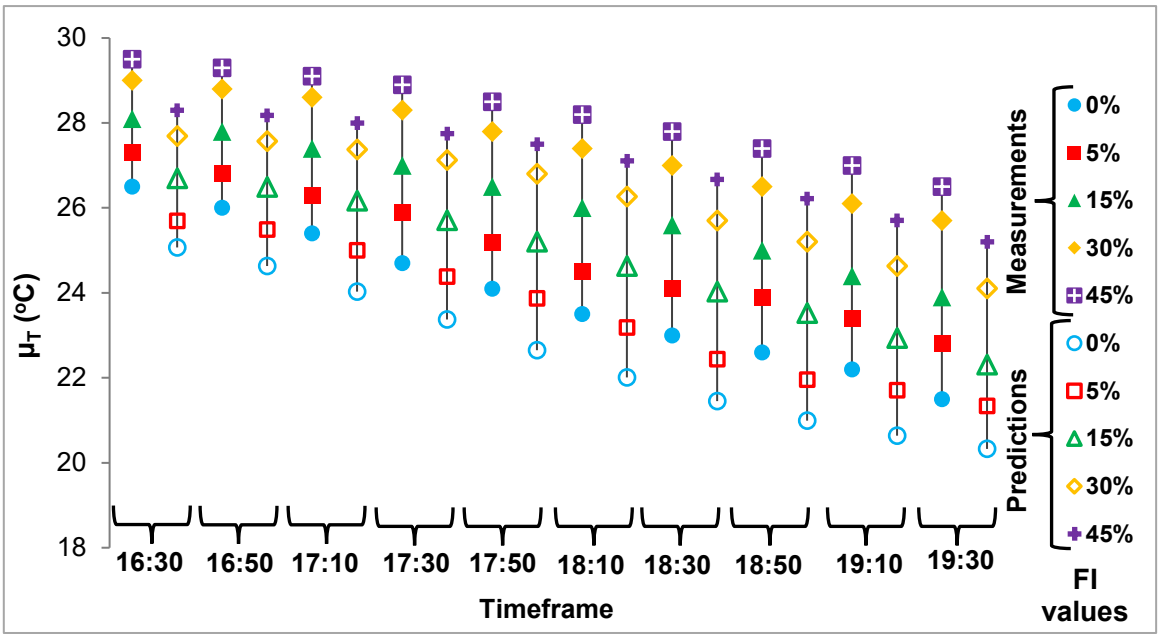
5.2.2 Projection of ballast surface temperature based on IRT images and LSTM predictions

- Mean surface temperature. Fig. 16 shows surface temperature of the ballast specimen including measured and LSTM-predicted values. Evidently, the distinct surface temperature patterns can be identified between characterized specimens with diverse FI percentages, as predictions derived from LSTM model strongly support this overall orientation via anticipating greater thermal values for highly contaminated ballast (FI value of 45%) regarding diurnal temperature fluctuations in the afternoon.

The variations of mean surface temperature derived for five diverse conditions of ballast fouling demonstrate that 5% and 15% FI values are close together, and the same trend is also observed in case of 30% and 45% FI values. Again, these variations confirm the superiority of defining three classes (instead of five classes) to estimate and classify ballast fouling conditions based on the IRT-captured surface temperatures. Accordingly, Fig. 17 compares median and quartile values of measured and predicted average surface temperatures for three distinct fouling levels over a seven-day period, in order to assess the accuracy of LSTM predictions against real measurements.



a Average ambient temperature: 36.9 °C



b Average ambient temperature: 23.8 °C

Fig. 16 Average surface temperature projections (measured and predicted) for fouled ballast specimens captured using IRT images and predicted via LSTM model

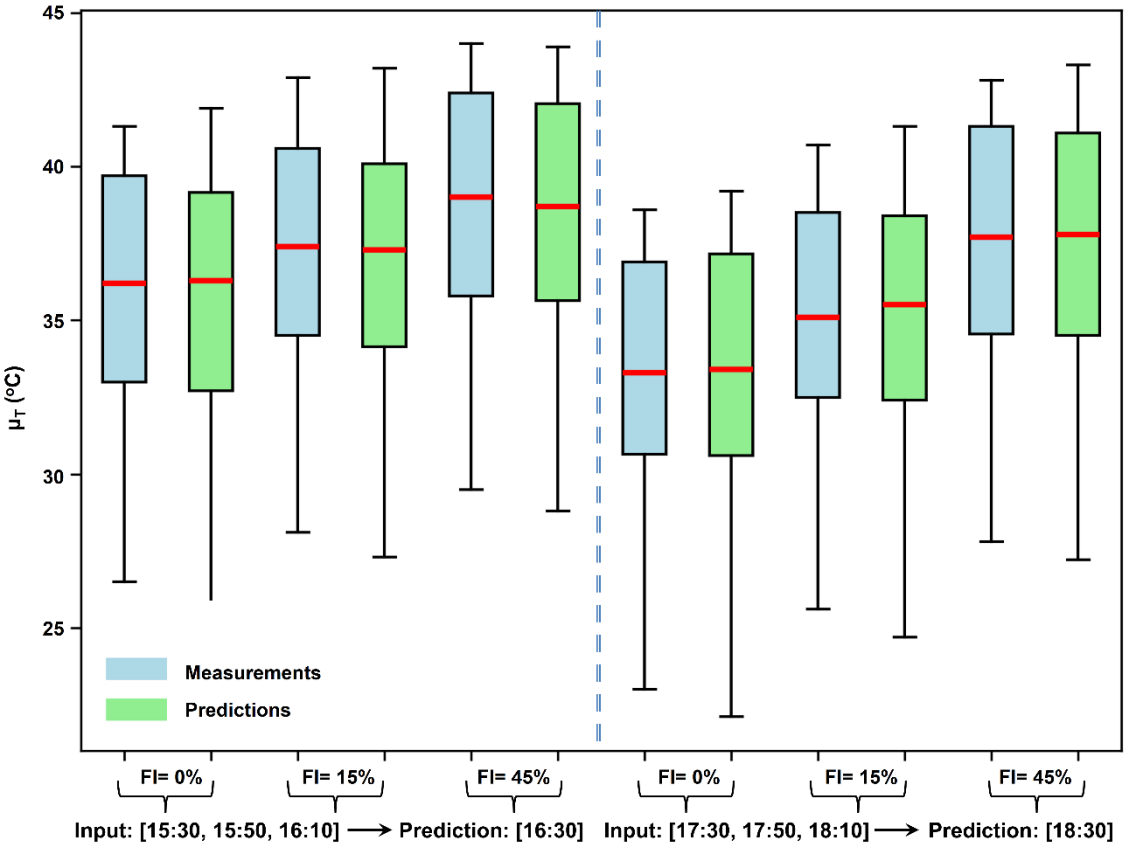
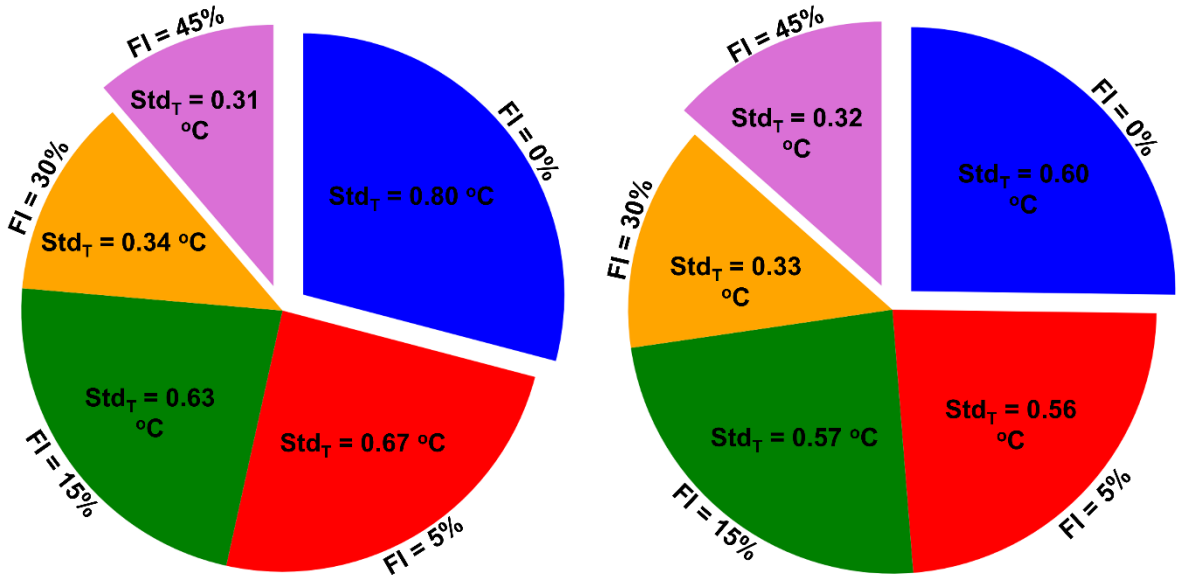


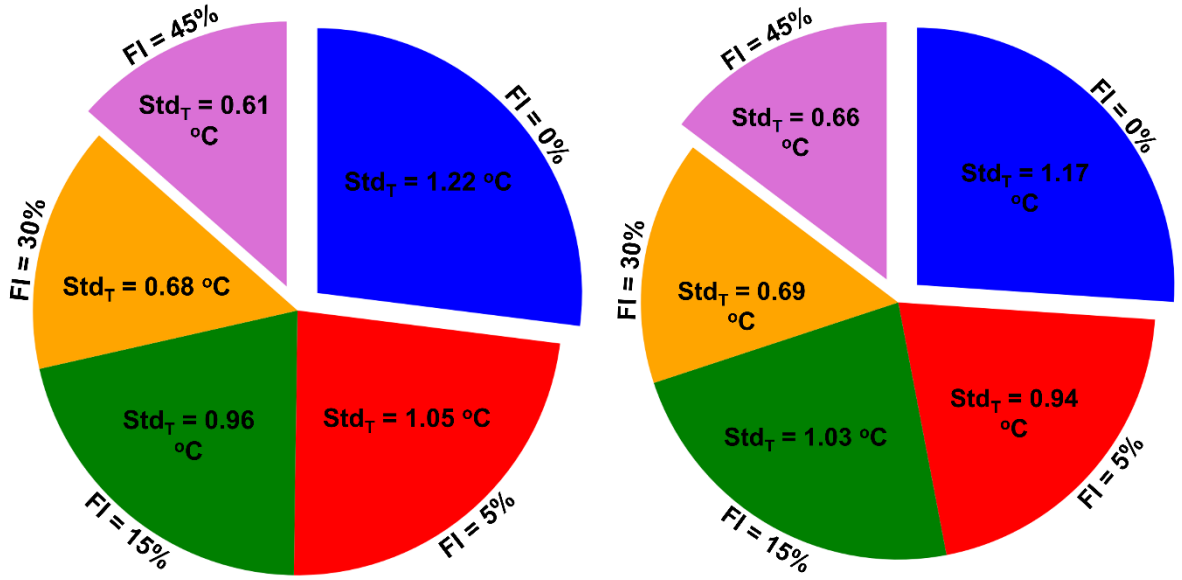
Fig. 17 Visual summary of measured and predicted average surface temperature projections for clean and fouled ballast specimens, demonstrating median and quartile values across seven consecutive days

- Standard deviation of surface temperature. To further assess the temperature distribution within the IRT images, Fig. 18 shows the standard deviation values of surface temperature (Std_T), both obtained and predicted, for characterized conditions of ballast fouling. Evidently, highly-contaminated ballast demonstrates more uniform variations of surface temperature, while clean specimens illustrate pronounced Std_T values due to their heterogeneous structure. Again, the predictions generated by the LSTM model provide robust support for this overarching trend, reinforcing the validity of the proposed framework.



a.1 Measured **a.1** Predicted

a Timeframe: 16:30 - 18:00



b.1 Measured **b.2** Predicted

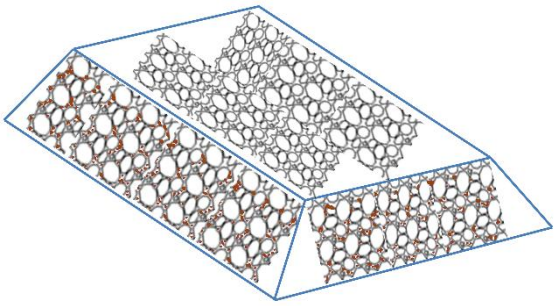
b 18:00 - 19:30

Fig. 18 Standard deviation of surface temperature captured using IRT images and predicted via LSTM model (Average ambient temperature: 36.9 °C)

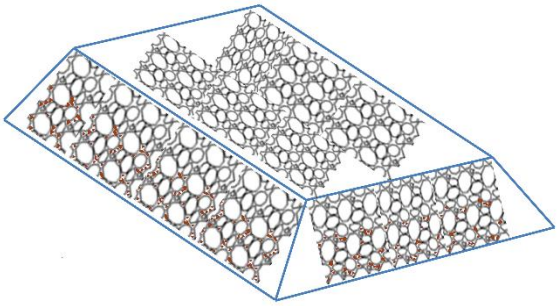
5.2.3 Robustness of LSTM-RF and LSTM-XGB models for ballast fouling classification

To assess the effectiveness of LSTM-RF and LSTM-XGB models developed for classification of ballast fouling conditions, real data of successive time-series surface temperatures measured through IRT are imported into the developed model. Three distinct fouling levels, including FI values of 0%, 10%, and 30%, along with two dispersion conditions of fouling agents among aggregates (uniform and non-uniform), are prepared. As illustrated in Fig. 19, the confusion matrix confirms the higher accuracy of predictions from LSTM-XGB to properly classify ballast conditions based on contamination. Moreover, the developed model demonstrates lower accuracy based on confusion matrix when the fouling materials are not evenly dispersed within the ballast aggregate. Thus, complementary NDT tools, such as GPR and ultrasound, are necessary to map surface and subsurface conditions. Overall, the robustness of detected conditions of ballast fouling is essential, because misclassification, whether estimating higher or lower levels of FI, can lead to the unnecessary implementation of maintenance activities or

extreme delay and potential speed restriction or downtime. As characterized by Chrismer and Hyslip (2018), exceeding FI value of 30% leads to lower effectiveness of ballast maintenance, mandating renewal of ballast through cleaning or replacement. Additionally, post-cleaning mechanical properties of ballast layer compromise operation safety, necessitating further train passage to restore its stability (Xiao et al. 2025b). Therefore, any misclassification with classifying ballast condition within higher class can lead to implementing unnecessary cleaning operation on railway tracks.

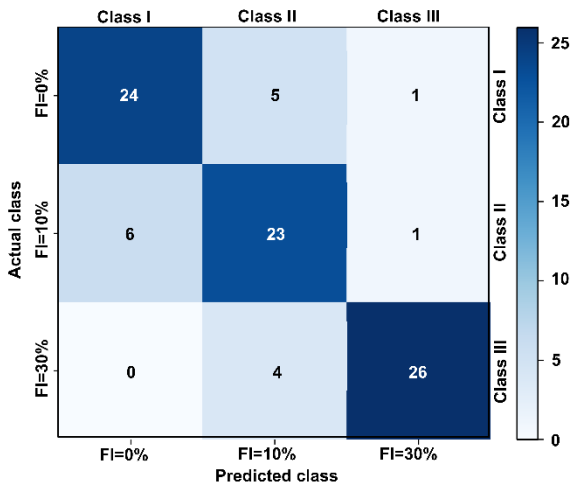


a.1 Uniform distribution of fouling material

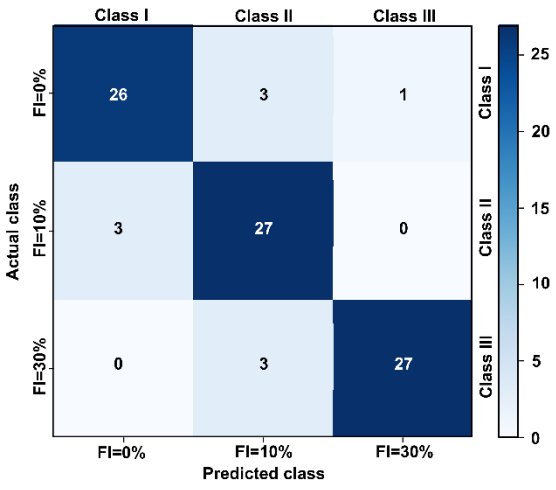


a.2 Non-uniform distribution of fouling material (contamination of lower thickness)

a Ballast layer

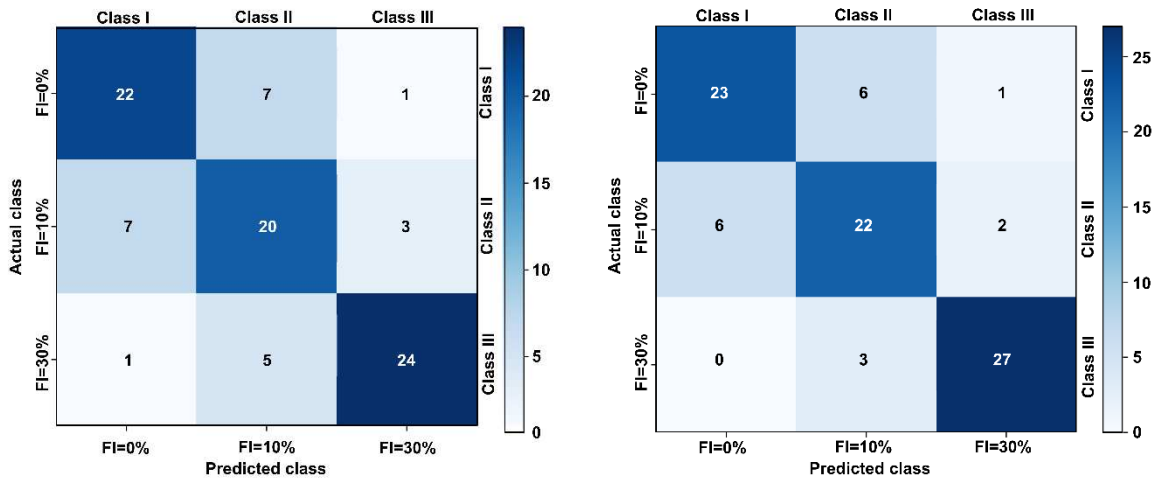


b.1 LSTM-RF classifier



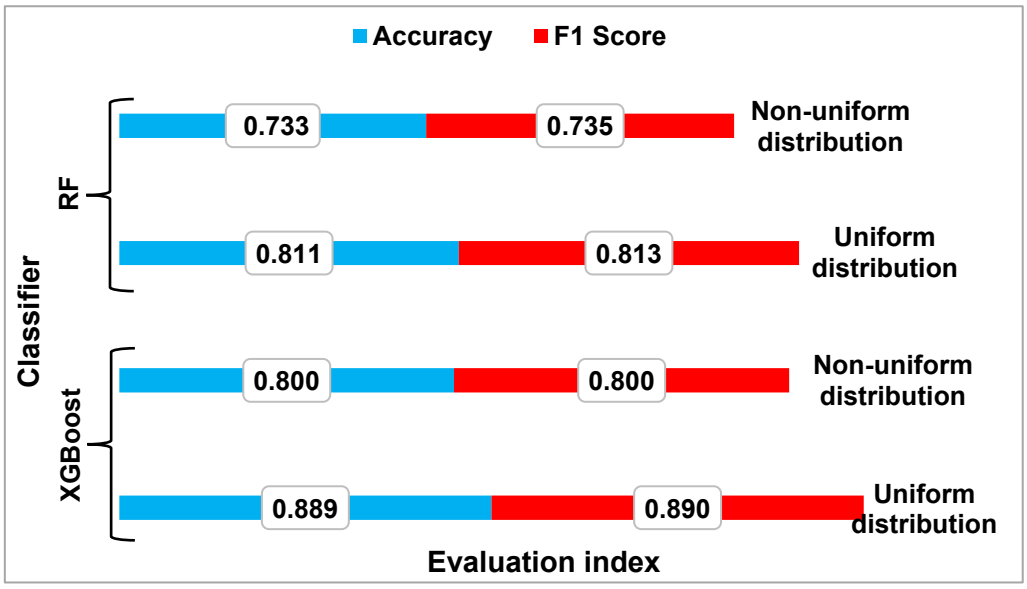
b.2 LSTM-XGBoost classifier

b Uniform distribution of clay in the ballast sample



c.1 LSTM-RF classifier **c.2** LSTM-XGBoost classifier

c Non-uniform distribution of clay in the ballast sample (contamination of lower thickness)



d Comparing values of accuracy and F1 Score derived based on developed classifier model

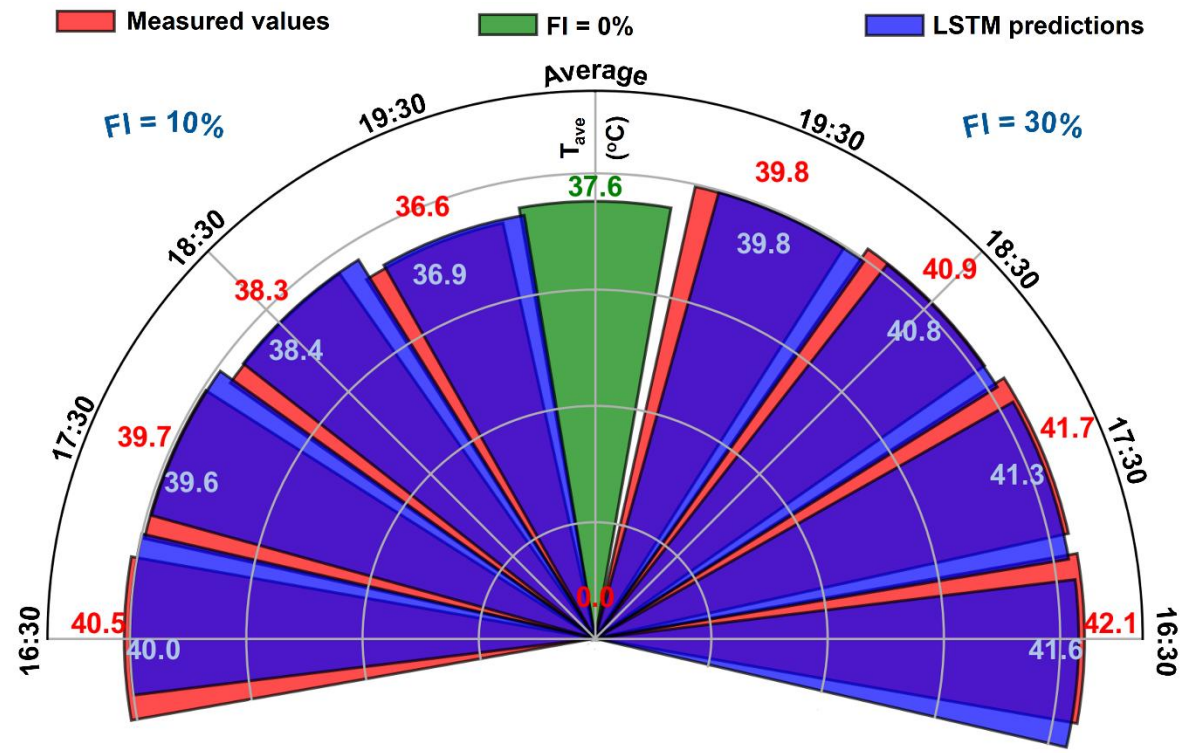
Fig. 19 Confusion matrix and evaluation indices demonstrating the robustness of developed model to classify fouled ballast for characterized conditions

5.3 Discussion

The LSTM-XGB model aids inspection operation in establishing a limited timeframe for the employment of IRT technology, while its predictive nature with respect to thermal time-series data captures upcoming temperatures. Moreover, the classification component automates the distinction of ballast contamination levels, ranging from clean to highly fouled specimens.

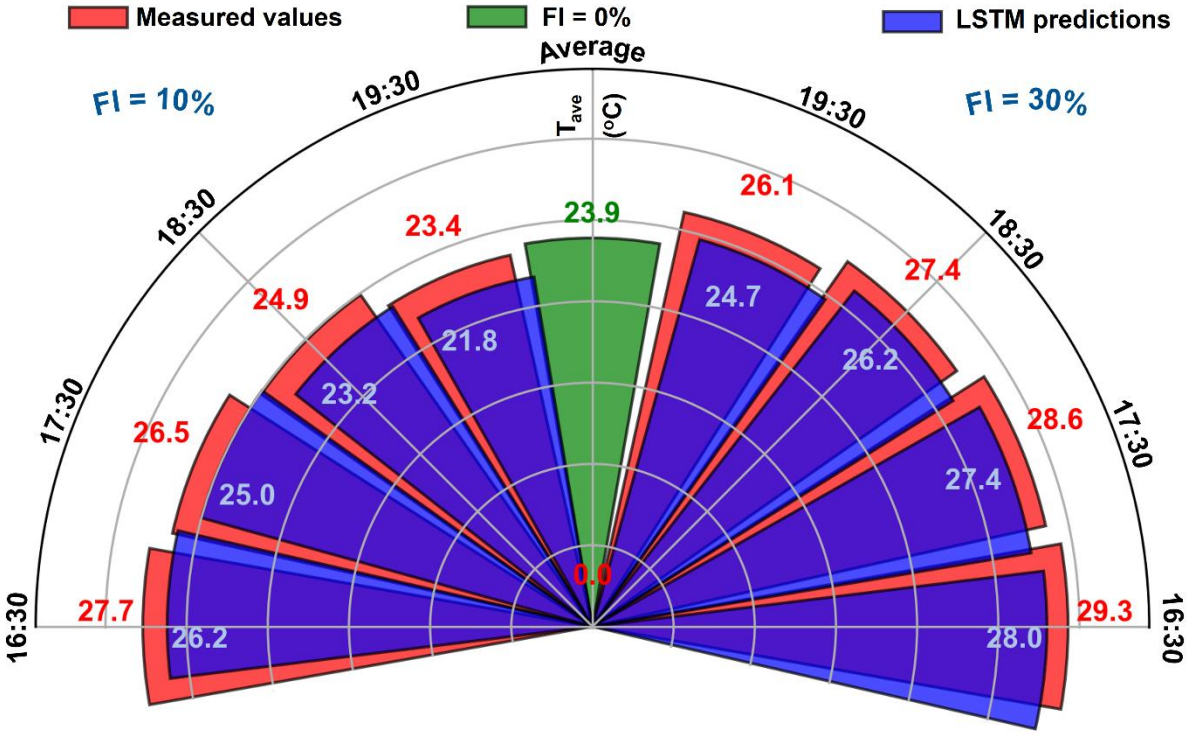
Regarding Fig. 20, the LSTM predictions comply well with the measured surface temperature values, capturing varying levels of FI. Likewise, as illustrated in Fig. 21, the integration of LSTM-XGB for classifying ballast fouling into three characterized classes demonstrates strong performance compared to the standalone deployment of the XGBoost model, enabling the use of three successive measured surface temperatures (rather than six measurements using IRT) for ballast fouling classification. Accordingly, IRT-derived FI classes support decision making for maintenance operations: $FI < 1\%$ represents clean ballast requiring routine monitoring; $1\% \leq FI < 20\%$ indicates moderately fouled ballast that warrants targeted inspection and preventive maintenance scheduling; and finally $FI \geq 20\%$ denotes heavily fouled ballast necessitating priority intervention trough ballast cleaning or renewal.

Overall, as characterized by Fayyad et al. (2025), unmanned aerial vehicles (UAVs) can be equipped with thermal imaging sensors to expedite the process of capturing thermal images from the ballast surface at specified time intervals, aiming to reveal variations in estimated surface temperature as an indicator of potential subsurface defects. Furthermore, the AI-integrated methodology improves the efficiency of SHM of the ballast layer based on IRT-captured thermal measurements. In this context, Hu et al. (2019) emphasized the automation of defect classification using AI algorithms, leading to a more reliable inspection system for ballasted railway tracks. Finally, the integration of machine vision technology (Xiao et al. 2025a) and GPR inspection enhanced by explainable artificial intelligence (Liu et al. 2023; Matsimbe et al. 2024) offers complementary tools that enable a comprehensive assessment of ballast conditions, non-destructive detection and segmentation of ballast fouling areas, and informed decision-making for railway track maintenance (e.g., cleaning operations).



525

526 **a** Average ambient temperature: 35.5 °C



527

528 **b** Average ambient temperature: 25.1 °C

529 **Fig. 20** Comparison of measured and LSTM predictions for surface temperature of ballast (capturing

530 FI values of 0%, 10%, and 30%)

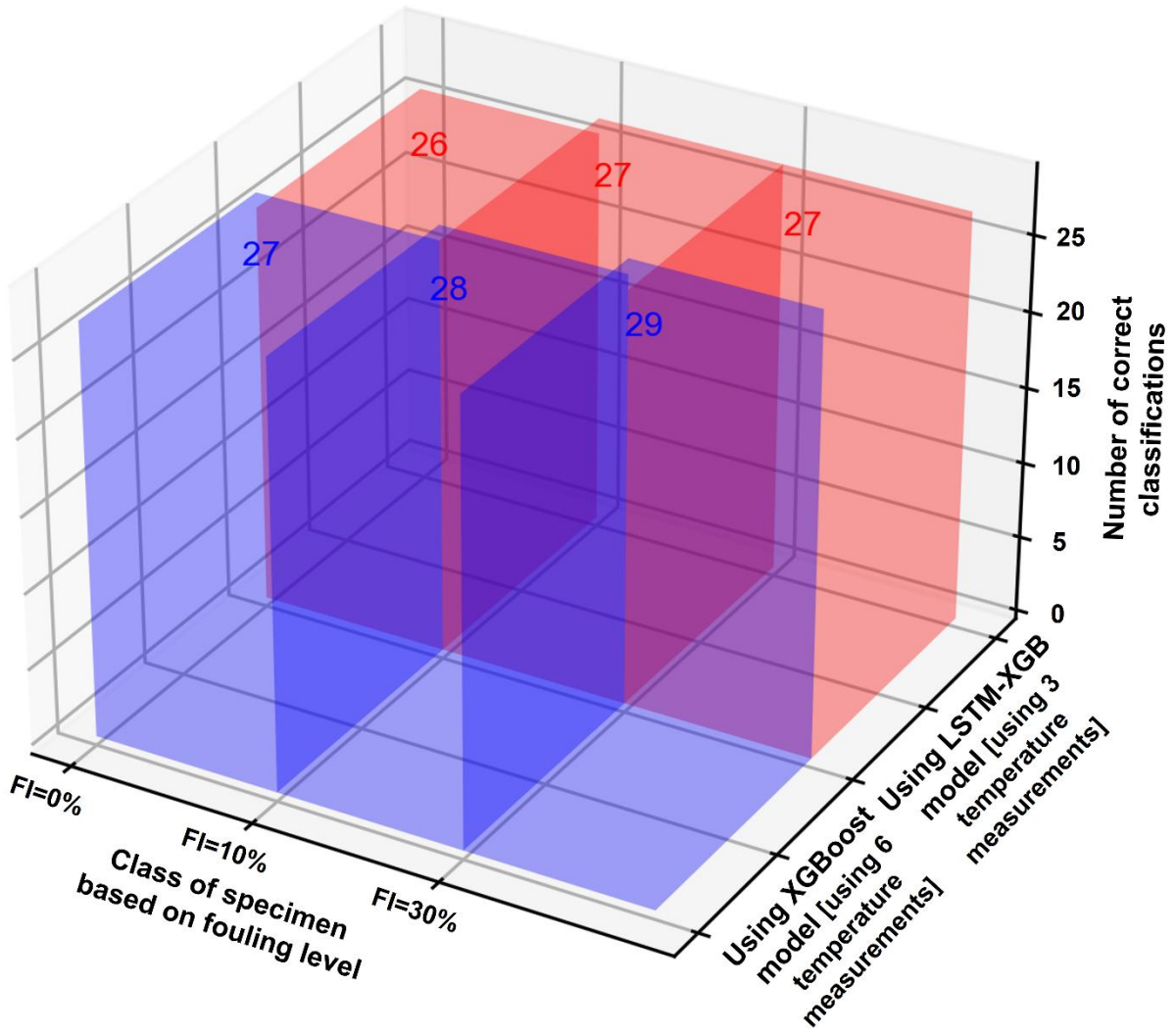


Fig. 21 Comparison of performance of XGBoost and LSTM-XGB models for classification of ballast fouling (capturing FI values of 0%, 10%, and 30%)

6 Conclusions and future perspectives

6.1 Conclusions

This study employed the developed LSTM-XGB model on IRT images captured from railway ballast surface to apply machine learning and deep learning techniques for non-destructive testing (NDT)-based monitoring. The fused model predicts surface temperature and classifies clay-fouled ballast conditions. The key findings are as follows:

1. An increase in fouling level leads to a reduction in cooling rates during diurnal temperature fluctuations in the afternoon, as higher surface temperatures are measured via IRT for heavily fouled ballast specimens. The LSTM predicts this general trend under diverse characterized conditions.

2. The temperature variations across thermogram pixels indicate that increasing fouling severity results in a reduced standard deviation of surface temperature.

3. The XGBoost model, compared to the RF model, when reinforced with measured time-series thermal data along with LSTM-predicted data, demonstrates a higher level of accuracy in classifying ballast according to contamination levels ranging from clean to highly-fouled specimens.

4. During the cooling down timeframe, higher contrast between clean and fouled ballast specimens is observed, as higher fouling index results in increased mean surface temperature and decreased standard deviation value. This trend is captured through the LSTM-XGB model to classify contamination level.

5. The time-series predictive nature of LSTM component of the LSTM-XGB model enables capturing shorter time window (three sequential IRT measurements rather than six) for measuring ballast surface temperature, leading to higher efficiency compared to the standalone deployment of the XGBoost model.

6. Overall, the two-stage integration of LSTM and XGBoost implemented on IRT data, capturing surface temperature variations of granular specimens, can assess ballast health with respect to the occurrence of fouling, particularly when the fouled-ballast condition is categorized into three groups instead of five (demonstrating a shift from finer to coarser classification).

6.2 Future perspectives

- The model developed based on the uniform dispersion of fouling materials among ballast particles demonstrates significantly lower accuracy for ballast specimens with fouling occurring within the lower part of the layer, thereby increasing the potential for misclassification. Therefore, future research should expand the database used for training ML/DL models to captures not only the occurrence of non-uniform fouling among ballast particles, but also the presence of other fouling materials, such as sand intrusion and aggregate degradation, as well as the presence of moisture among fine contents.

- Using active excitations (such as halogen or IR lamps) would be complementary to enhance contrast particularly during winter periods as well as strengthening the robustness of the method.

- The developed AI-enabled passive IRT method should be integrated into full-scale field applications under uncertain environmental conditions, such as wind, solar radiation, and moisture retention or transition, to validate the robustness of the approach in operating environments.

- Fusing IRT with complementary inspection technologies capable of assessing subsurface conditions (such as GPR) offers a practical pathway to enhance diagnostic efficiency. Moreover, to assess the practical usefulness of the established method, the thermal imaging system could be mounted on a moving cart or an autonomous drone to continuously monitor ballast health with respect to fouling, while also accounting for field challenges such as motion blur, standoff distance, and shading from sleepers, rails, or the moving platform, which can affect the reliability of fouling detection by blurring thermal features, altering emissivity accuracy, or distorting the geometry of the region of interest.

Compliance with ethical standards

The authors declare that there is no conflict of interest regarding the publication of this article.

References

Adeagbo, M. O., Wang, S. M., & Ni, Y. Q. (2024). Revamping structural health monitoring of advanced rail transit systems: A paradigmatic shift from digital shadows to digital twins. *Advanced Engineering Informatics*, *61*, 102450.

Aela, P., Cai, J., Jing, G., & Chi, H. L. (2024). Vision-based monitoring of railway superstructure: A review. *Construction and Building Materials*, *442*, 137385.

AlDousari, A. E., Fattah, M. A., & Kafy, A. A. (2024). Assessing Solar Irradiance Trends and Temperature Extremes by Applying Machine Learning Based Statistical Modelling for Renewable Energy Optimization. *Earth Systems and Environment*, *8*(4), 1523-1539.

Alsahli, A., & Alsulmi, M. (2023). Automatic detection of sand fouling levels in railway tracks using supervised machine learning: a case study from Saudi Arabian railway. *Arabian Journal for Science and Engineering*, *48*(4), 4925-4935.

Anglade, E., Aubert, J. E., Sellier, A., & Papon, A. (2022). Physical and mechanical properties of clay–sand mixes to assess the performance of earth construction materials. *Journal of Building Engineering*, *51*, 104229.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Chapman, L., Thornes, J. E., & White, S. P. (2006). Thermal imaging of railways to identify track sections prone to buckling. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *220*(3), 317-327.

Chen, L., Chen, J., Wang, C., Dai, Y., Guo, R., & Huang, Q. (2021). Modeling of Moisture Content of Subgrade Materials in High-Speed Railway Using a Deep Learning Method. *Advances in Materials Science and Engineering*, *2021*(1), 6166489.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Chrismer, S., & Hyslip, J. (2018). Principles of Degraded Ballast and Their Track Safety Implications. In *AREMA Annual Conference*.

Cui, S., Sun, X., Luan, M., & Wei, Y. (2021). Research on crack detection method of ballastless track slab based on infrared thermometer. *Infrared Physics & Technology*, *116*, 103772.

D'Accardi, E., Dell'Avvocato, G., Masciopinto, G., Marinelli, G., Fumarola, G., Palumbo, D., & Galietti, U. (2025). Evaluation of typical rail defects by induction thermography: experimental results and procedure for data analysis during high-speed laboratory testing. *Quantitative InfraRed Thermography Journal*, *22*(2), 173-194.

De Simone, L., Caputo, E., Cinque, M., Galli, A., Moscato, V., Russo, S., ... & Giannini, G. (2023). LSTM-based failure prediction for railway rolling stock equipment. *Expert Systems with Applications*, *222*, 119767.

Deng, Z., He, Y., Li, Y., Xu, L., Xiao, Y., & Su, Q. (2025). Monitoring and prediction of surface deformation along railways lines in karst areas using multi-source data− a case study of the Beijing-Guangzhou Railway in China. *Transportation Geotechnics*, 101564.

Farhadi, H., & Najafzadeh, M. (2021). Flood risk mapping by remote sensing data and random forest technique. *Water*, *13*(21), 3115.

Fayyad, T. M., Taylor, S., Feng, K., & Hui, F. K. P. (2025). A scientometric analysis of drone-based structural health monitoring and new technologies. *Advances in Structural Engineering*, *28*(1), 122-144.

FLIR (2014). User's manual, Publ. No.: T559795 in English.

Gers, F. A., & Schmidhuber, E. (2001). LSTM recurrent networks learn simple context-free and context-sensitive languages. *IEEE transactions on neural networks*, *12*(6), 1333-1340.

Guclu, A., Yılboga, H., Eker, O. F., Camci, F., & Jennions, I. (2010, October). Prognostics with autoregressive moving average for railway turnouts. In *Annual Conference of the PHM Society* (Vol. 2, No. 1).

He, Y., Deng, B., Wang, H., Cheng, L., Zhou, K., Cai, S., & Ciampa, F. (2021). Infrared machine vision and infrared thermography with deep learning: A review. *Infrared physics & technology*, *116*, 103754.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Hu, C., Duan, Y., Liu, S., Yan, Y., Tao, N., Osman, A., ... & Zhang, C. (2019). LSTM-RNN-based defect classification in honeycomb structures using infrared thermography. *Infrared Physics & Technology*, *102*, 103032.

Hussaine, S. M., Mu, L., Lu, Y., & Hussain, S. S. (2025). Forecasting Tunnel-Induced Ground Settlement: A Hybrid Deep Learning Approach and Traditional Statistical Techniques With Sensor Data. *IEEE Access*.

Guo, Y., Xie, J., Fan, Z., Markine, V., Connolly, D. P., & Jing, G. (2022). Railway ballast material selection and evaluation: A review. *Construction and Building Materials*, *344*, 128218.

Indraratna, B., Su, L. J., & Rujikiatkamjorn, C. (2011). A new parameter for classification and evaluation of railway ballast fouling. *Canadian Geotechnical Journal*, *48*(2), 322-326.

Jang, B. S., Kim, Y., Kim, S., Choi, H. J., & Yoon, H. K. (2023). Assessment of Applicability of CNN Algorithm for Interpretation of Thermal Images Acquired in Superficial Defect Inspection Zones. *Journal of The Korean Geotechnical Society*, *39*(10), 41-48.

Jang, B. S., & Yoon, H. K. (2024). Application of infrared thermal images for sinkhole detection with time-series and time-difference index data through a convolution neural network. *International Journal of Pavement Engineering*, *25*(1), 2317432.

Kapoor, R., Goel, R., & Sharma, A. (2020). Deep learning based object and railway track recognition using train mounted thermal imaging system. *Journal of Computational and Theoretical Nanoscience*, *17*(11), 5062-5071.

Koohmishi, M., Kaewunruen, S., Jing, G., & Guo, Y. (2024). Thermal imaging analysis of ballast fouling: Investigating the effects of parent rock and fouling materials through IRT passive camera. *Transportation Geotechnics*, *48*, 101313.

Koohmishi, M., Gao, Y., Jing, G., & Guo, Y. (2025). Non-destructive assessment of railway ballast fouling and water retention using infrared thermography and statistical processing. *Construction and Building Materials*, *466*, 140273.

Lai, Y., & Dzombak, D. A. (2020). Use of the autoregressive integrated moving average (ARIMA) model to forecast near-term regional temperature and precipitation. *Weather and forecasting*, *35*(3), 959-976.

Li, Z. W., Liu, X. Z., Lu, H. Y., He, Y. L., & Zhou, Y. L. (2020). Surface crack detection in precasted slab track in high-speed rail via infrared thermography. *Materials*, *13*(21), 4837.

Liang, X., Niu, X., Liu, P., Lan, C., Yang, R., & Zhou, Z. (2023). Test on fouling detection of ballast based on infrared thermography. *NDT & E International*, *140*, 102956.

Liang, X., Yang, R., Zhu, W., Lan, C., Ding, X., Zhang, Q., ... & Ding, H. (2024a). Observing and identifying fouled ballast bed: On-site testing with infrared thermography (IRT) and uncovering thermodynamic transfer mechanisms within the ballast bed. *NDT & E International*, *147*, 103168.

Liang, X., Yang, R., Liu, P., Niu, X., Zhang, Q., Ding, H., ... & Chen, J. (2024b). Advancing railway infrastructure maintenance: Thermodynamic parameter inversion of ballast bed and feasibility assessment of fouling detection via infrared thermography (IRT). *Infrared Physics & Technology*, 105398.

Liang, X., Yang, R., Qian, H., Yang, Z., Zhang, Q., Geng, H., ... & Chen, J. (2025). Observing and identifying fouled ballast bed using infrared thermography (IRT): A real-time temperature prediction study based on an enhanced BiGRU model. *Mechanical Systems and Signal Processing*, *224*, 112150.

Liu, G., Yang, F., Wang, S., Jing, G., & Nateghi, Y. (2023). Railway ballast fouling, inspection, and solutions-A review. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *237*(8), 969-982.

Loche, M., Scaringi, G., Blahůt, J., & Hartvich, F. (2022). Investigating the potential of infrared thermography to inform on physical and mechanical properties of soils for geotechnical engineering. *Remote Sensing*, *14*(16), 4067.

Matsimbe, F. N. B., Michael, A., & Choi, Y. T. (2024). Explainable Machine Learning Model for Ballast Condition Assessment Using Ground Penetrating Radar Scans. In *International Conference on Civil, Structural and Transportation Engineering*. Avestia Publishing.

Mishra, M., Odelius, J., Thaduri, A., Nissen, A., & Rantatalo, M. (2017). Particle filter-based prognostic approach for railway track geometry. *Mechanical systems and signal processing*, *96*, 226-238.

Pedram, M., Taylor, S., Hamill, G., Robinson, D., OBrien, E. J., & Uddin, N. (2022). Experimental evaluation of heat transition mechanism in concrete with subsurface defects using infrared thermography. *Construction and Building materials*, *360*, 129531.

Pozzer, S., Omidi, Z., El Refai, A., López, F., Ibarra-Castanedo, C., & Maldague, X. (2024). Passive infrared thermography for subsurface delamination detection in concrete infrastructure: Capabilities. *Construction and Building Materials*, *419*, 135542.

Ramadan, A. N., Zhang, J., Jing, P., Zhang, L., & Murtaza, M. (2025). Evaluating the hydraulic behavior of sandy fouled ballast: A case study from Inner Mongolia, China. *Transportation Geotechnics*, *50*, 101437.

Ramzan, B., Malik, M. S., Martarelli, M., Ali, H. T., Yusuf, M., & Ahmad, S. M. (2021). Pixel frequency based railroad surface flaw detection using active infrared thermography for structural health monitoring. *Case Studies in Thermal Engineering*, *27*, 101234.

Rana, A. (2022). *Experimental Testing and Analysis of Moving Infrared Thermography for Rail Base Inspection* (Master's thesis, Southern Illinois University at Carbondale).

Rojas-Vivanco, J., Breul, P., Talon, A., García, J., Navarrete, M. B., & Villavicencio, G. (2025). Assessment of railway ballast fouling using GPR and AI-Based learning from LDCP and geoendoscopy data. *Transportation Geotechnics*, 101701.

Schnebele, E., Tanyu, B. F., Cervone, G., & Waters, N. J. E. T. R. R. (2015). Review of remote sensing methodologies for pavement management and assessment. *European Transport Research Review*, *7*(2), 7.

Selig, E. T., & Waters, J. M. (1994). *Track geotechnology and substructure management*. Thomas Telford.

Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). *Time series analysis and its applications* (Vol. 3, p. 4). New York: springer.

Sohn, H., Farrar, C. R., Hemez, F. M., Shunk, D. D., Stinemates, D. W., Nadler, B. R., & Czarnecki, J. J. (2003). A review of structural health monitoring literature: 1996–2001. *Los Alamos National Laboratory, USA*, *1*(16), 10-12989.

Wakjira, T. G., Ibrahim, M., Ebead, U., & Alam, M. S. (2022). Explainable machine learning model and reliability analysis for flexural capacity prediction of RC beams strengthened in flexure with FRCM. *Engineering Structures*, *255*, 113903.

Wang, X., Bai, Y., & Liu, X. (2023). Prediction of railroad track geometry change using a hybrid CNN-LSTM spatial-temporal model. *Advanced Engineering Informatics*, *58*, 102235.

Wongkaew, W., Muanyoksakul, W., Ngamkhanong, C., Sresakoolchai, J., & Kaewunruen, S. (2024). Data driven machine learning prognostics of buckling failure modes in ballasted railway track. *Discover Applied Sciences*, *6*(4), 212.

Xiao, C., Chen, N., Hu, C., Wang, K., Gong, J., & Chen, Z. (2019). Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. *Remote Sensing of Environment*, *233*, 111358.

Xiao, Y., Ning, Y., Peng, Y., Wang, M., Long, Y., Tan, S., ... & Yu, Z. (2025a). A novel machine-vision-based algorithm for quantifying surface fouling of railway ballast beds. *Advanced Engineering Informatics*, *68*, 103639.

Xiao, Y., Shi, S., Xia, M., Wang, J., Niu, Z., Wang, C., & Gao, L. (2025b). Failure analysis and improvement of trackbed mechanical properties in ballast cleaning operation. *Engineering Failure Analysis*, 109771.

Xu, L., & Hu, J. (2021). A method of defect depth recognition in active infrared thermography based on GRU networks. *Applied Sciences*, *11*(14), 6387.

Yang, X., Huang, R., Meng, Y., Liang, J., Rong, H., Liu, Y., ... & Feng, Y. (2024). Overview of the Application of Ground-Penetrating Radar, Laser, Infrared Thermal Imaging, and Ultrasonic in Nondestructive Testing of Road Surface. *Measurement*, 113927.

Ye, X. Y., Luo, Y. Y., Li, Z. W., & Liu, X. Z. (2023). A Quantitative Detection Method for Surface Cracks on Slab Track Based on Infrared Thermography. *Applied Sciences*, *13*(11), 6681.

Yürekli, Y., Özarpa, C., & Avcı, İ. (2025). Real-Time Railway Hazard Detection Using Distributed Acoustic Sensing and Hybrid Ensemble Learning. *Sensors*, *25*(13), 3992.