Data Engineering and Applications

Research Paper

Concrete Pouring Quality Assessment S4D-Transformer Model Driven by Thermal Video and Pump Pressure Data

Sultan Al-Dhaheri[1,*], Abdullah Al-Zaabi[1] and Juma Al-Sharqi[1]

[1] School of Engineering, Computer Science Department, Khalifa University, Abu Dhabi, PO Box 127788, United Arab Emirates
*Corresponding author: sultan.dh@kustar.ac.ae

Abstract. This study develops an S4D-Transformer model for concrete pouring quality assessment driven by thermal video and pump-pressure data. The task aims to detect cold-joint risk, segregation, inadequate vibration, pressure interruption, unstable pumping, and abnormal hydration-temperature response before defects become difficult to correct after hardening. A simulated construction dataset is built with 9,800 pouring windows, 420 thermal video sequences, 18 pump-pressure channels, 36 structural zones, and four member types: wall, slab, column, and beam sections. The proposed workflow aligns thermal frames and pressure pulses in fixed pouring windows, extracts surface-temperature texture, pressure fluctuation, flow stability, pumping pause, vibration-response, pouring-route, and zone-context features, and then combines S4D long-range state memory with Transformer cross-modal fusion. Compared with pressure-only, thermal CNN, LSTM fusion, temporal convolution, and Vision Transformer baselines, the model improves pouring-quality classification accuracy to 92.8%, raises cold-joint recall to 90.7%, and reduces false alarms by 26.3%. Average inference delay is 41 ms per window, supporting near-real-time site review. The method should be interpreted as an engineering decision-support tool and still requires validation with real thermal cameras, pump-controller logs, vibration records, field inspection labels, and project-specific acceptance criteria.

Keywords: *Thermal Video Assessment, Pump Pressure Sequence, Concrete Pouring Quality, S4D-Transformer, Cold Joint Detection, Construction Process Monitoring*

Received on 26 October 2025, Accepted on 02 May 2026, Published on 12 May 2026

Introduction

Fluctuations in temperature, pump pressure changes, vibration status and simultaneous surface texture variations are all problems for real-time quality inspection during concrete pouring [1]. Thermal imaging has a small temperature difference, and thermal contrast is also affected by weather, formwork shadows, material thickness, curing conditions, and camera angle [2]. Pump-pressure data can show a delivery interruption, a trend of pipe obstruction and unstable flow; however, a pressure peak is not always caused by a serious quality defect [3]. Deep sequence models can connect multiple construction data streams over time in the absence of manual observation [4]. The foreman will observe the speed of pouring, surface smoothness, bleeding, vibration marks and placement continuity at the construction site. These cues are convenient but subjective, and they are less so when several pouring zones work together. A digital model can organise thermal video and pressure data into a continuous-quality record.

When concrete is poured in areas with dense reinforcement, long wall sections, high columns, large slabs and beam-column connections, multi-source inspection is required. Construction monitoring studies [5] have shown that variable inspection standards, a lack of supervision records and irregular data are common problems in site quality documentation. Research on thermal monitoring [6] links the early-stage distribution of temperature with moisture and setting behaviour, and pressure-based process diagnosis [7] relates pumping stability to flowability and delivery resistance. The two data sources may be different. A stable pressure curve will still be observed if the local vibration is not severe, and the thermal pattern will be reasonable during an intermediate delivery interruption. Research on intelligent construction supports integrated sensing [8], but the actual system still needs to address long sequences, short interruptions and cross-zone variations.

The three kinds of models are: First, it should retain long-range sequence information because cold-joint risk may be caused by several earlier pauses rather than one anomalous frame. Second, it should preserve local events as a vibration mark, heat streak or short pressure pulse can be an important quality cue. Third, the output should indicate which zone, time window and signal source triggered the alarm for the site engineer to verify. Transformer-based fusion [9] can connect thermal patterns and pressure states, and state-space sequence modeling [10] is suitable for long-range dependencies. The proposed S4D-Transformer combines the above strengths to process thermal video and pump-pressure data simultaneously and to generate quality scores for cold-joint risk, segregation, inadequate vibration, and unstable pumping.

This paper's structure is not predicated on a particular method, but rather on the construction quality evaluation objective. In Section 2, sequence modelling, pump pressure signal analysis, and thermal video monitoring are used to evaluate the quality of construction. The synchronous thermal-pressure data stream, the S4D-Transformer design, the quality-scoring goal, and the warning output are all introduced in Section 3. Using data figures and a comparison table, Section 4 examines the simulated dataset, prediction accuracy, defect reaction, ablation behaviour, robustness, latency, and practical deployment. The primary conclusion and shortcomings of the approach are presented in Section 5.

Related Work

Thermal Video-Based Concrete Construction Monitoring

Thermal imaging is used during construction inspections to detect temperature variations that are difficult to see in a standard RGB video. Concrete work's surface temperature can reveal information about formwork contact, material discontinuity, hydration reaction, surface moisture, and environmental exposure. Zhu and Wang [11] showed that early-age concrete temperature evolution is closely related to heat-transfer conditions and crack-control behavior, which supports the use of thermal change as process evidence rather than a simple surface image. Guo et al. [12] further indicated that image-based concrete defect assessment can capture weak visual abnormalities under low-light, overexposed, and blurred conditions, and this is relevant to thermal inspection because construction-site images often contain incomplete and unstable visual cues. However, there is no direct connection between the internal features and thermal pictures. Solar radiation, camera distance, wind, and curing conditions all affect temperature patterns. Thermal evidence and contextual process data have recently been combined in vision-based construction monitoring, and Kang et al. [13] showed that concrete defect detection benefits from linking image evidence with practical deterioration categories. It refers to the combination of zone-level temperature differential, surface cooling rate, and thermal texture interpretations under pumping and vibration circumstances for pouring quality evaluation.

Pump Pressure Signal Analysis and Pouring Process Diagnosis

An indirect process indication of the delivery of ready-mixed concrete is pump pressure. Keep track of pressure pulses brought on by local obstructions, delivery height, cylinder switching, pipe friction, and mixture consistency. Construction equipment signals can anticipate process instability prior to the final product inspection, according to studies on pressure monitoring [14]. Additionally, Concrete Pumping Analysis [15] demonstrates the relationship between a change in pressure and variations in workability and flow resistance. Pump pressure, on the other hand, is a local or semi-local indication. It may discover that the delivery method is erratic, but it is not always able to identify the specific structural section that has a quality issue. In order to correlate local visual evidence with pump-side recordings, sensor fusion research [16] is required. In this article, pressure signals are not considered separate time series. They are mapped to thermal windows after being transformed into pulse characteristics, pause length, pressure-slope descriptors, and flow-stability indicators.

Sequence Modeling for Construction Quality Assessment

Extended time series and local disruptions are common in construction quality assessments. Sequence dependency may be addressed by Recurrent Neural Networks (RNNs), and while temporal convolution works well for analysing local patterns, both are inadequate for cross-modal reasoning and stable long-term memory. Because attention may connect distant regions, transformer models [17] have been widely used in sequence fusion. S4-style models [18] have demonstrated strong performance in learning long-range sequences, while state-space models [19] are another approach that models lengthy sequences with an effective recurrence-like structure. State-space memory and attention have just started to be included into hybrid sequence models [20]. This is how the suggested model proceeds. Transformer attention is utilised to link pressure changes with changes in thermal texture within the same structural area, while S4D is used to maintain the long-pouring-state development.

Methodology

Multi-Source Pouring Data Construction

The input is constructed using thermal video frames and pump-pressure data from a simulated concrete-pouring operation. 120 thermal frames sampled at 10Hz are included in each 12-second pouring window; pump pressure is resampled at the same time grid. Four types of members—walls, slabs, columns, and beams—and 36 pouring zones make up the simulated project. A pouring path, pump outlet distance, reinforcement density, ambient temperature, and vibration data are all linked to each zone. The entire process from thermal-pressure collection to quality evaluation output is depicted. The data flow, preprocessing path, feature building, synchronisation rule, and warning output may all be seen as a single construction process if Figure 1 is positioned before the model structure.

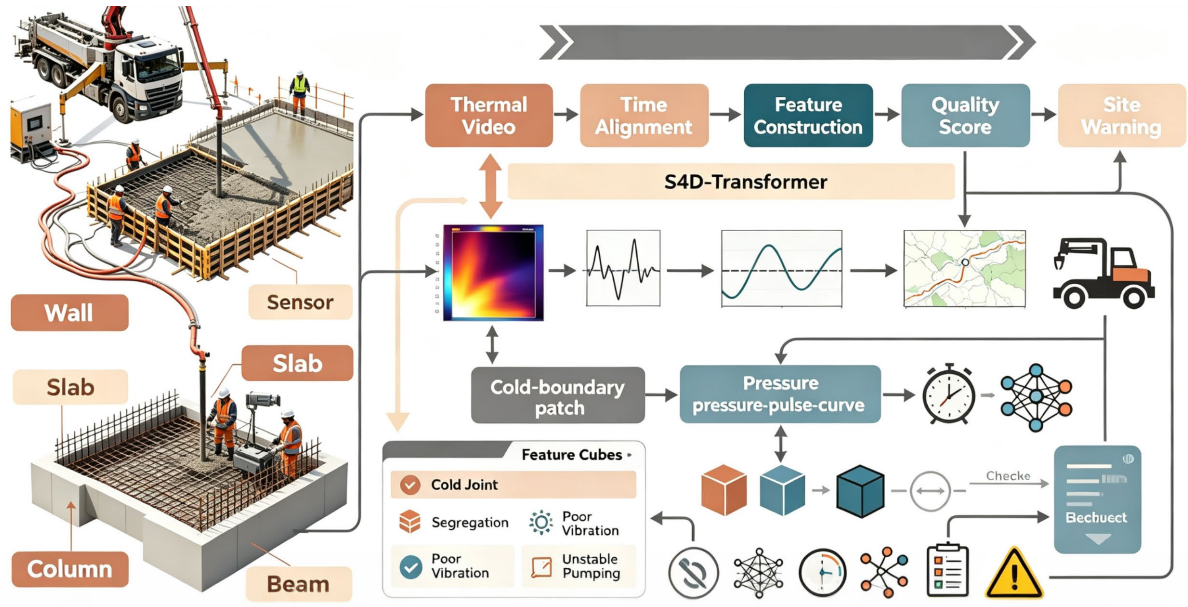

Figure 1. Thermal-pressure concrete pouring quality assessment workflow.

The thermal video sequence is first converted into a zone-level temperature field. The notation uses a normalized frame because the model should compare temperature evolution across structural members rather than memorize absolute camera brightness.

$$T_t = \mathrm{norm}\left(\mathrm{frame}_t^{\mathrm{thermal}}, A_z, H_z \right) \qquad \text{Eq.(1)}$$

The pump pressure sequence is aligned with the thermal window by timestamp and construction zone. Pressure values are represented with local slope and pulse amplitude so that short delivery disturbances remain visible after resampling.

$$P_t = [p_t, p_t - p_{t-1}, \max(p_{t-k:t}) - \min(p_{t-k:t})] \quad \text{Eq.(2)}$$

The multi-source feature vector combines thermal texture, pressure fluctuation, pouring-route position, reinforcement density, vibration state, and elapsed time from the last pouring pause. In the simulated dataset, each time step contains 46 normalized dimensions, including 18 thermal descriptors, 12 pressure descriptors, and 16 construction-context descriptors.

$$x_t = \mathrm{concat}(T_t, P_t, z_t, v_t, c_t) \quad \text{Eq.(3)}$$

False quality tags won't be produced by the synchronisation rule. The window will be retained for robustness testing but removed from the primary training if the temperature and pressure timestamps are greater than 80 ms apart. Each window has been categorised as qualified pouring, cold-joint danger, segregation risk, insufficient vibration, or unstable pumping based on simulated process records and visual evidence. There isn't just one criterion that determines the labels. A segregation label has an uneven surface texture and pressure instability; a cold-joint label has a pause length, thermal boundary evidence, and zone transition; and an inadequate-vibration label displays a poor vibration response and local thermal unevenness. Label construction is therefore more in line with the engineering site evaluation.

S4D-Transformer Feature Encoding

There are two collaborating streams in the S4D-Transformer concept. Transformer attention is used for inter-window linkages between heat and pressure evidence, whereas S4D sequence encoding is used in one stream to preserve long-range pouring-state memory. The S4D-Transformer feature encoding and quality rating system is shown in Figure 2. Thermal patches, pressure pulses, zone context, and vibration indicators all use different embedding layers. Since the S4D block simulates the ongoing pouring-state development, subsequent windows will be impacted by earlier pauses, pressure buildup, or temperature accumulation. A Transformer block with cross-modally fused features assesses if the pressure pulse and temperature boundary call for a warning before the quality head generates a danger score.

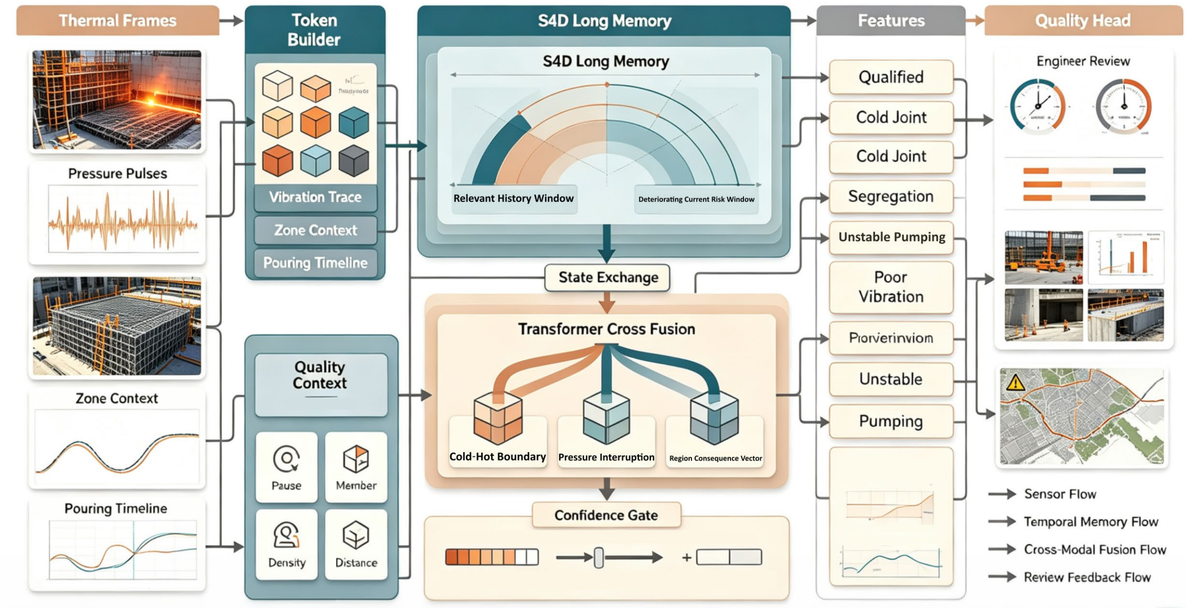

Figure 2. S4D-Transformer feature encoding and quality scoring structure.

The embedded input sequence is mapped into a hidden representation. This step keeps the time order of the pouring process and allows different sensor sources to share a common feature space.

$$e_t = W_e x_t + b_e \quad \text{Eq.(4)}$$

The S4D state update is used to preserve long-range pouring information. The state records whether a current quality response is linked to earlier pauses, pressure build-up, or thermal accumulation across the same zone.

$$s_t = A_d s_{t-1} + B_d e_t \quad \text{Eq.(5)}$$

The observable S4D feature is produced from the hidden state and the current embedded input. This feature describes temporal continuity without requiring full attention over every previous frame, which is useful when a pumping pause influences the thermal response several windows later.

Transformer fusion is then applied to connect thermal and pressure tokens. The attention coefficient becomes larger when a pressure pulse and a thermal boundary appear in related parts of the same pouring window.

$$a_{i,j} = \mathrm{softmax}\left(\frac{Q_i K_j^T}{\sqrt{d}}\right) \quad \text{Eq.(6)}$$

After obtaining the fused representation by combining the S4D feature with the cross-modal attention output, a lightweight feed-forward layer is applied. This design retains both local abnormal events and slow pouring-state development. Because the model must perform near-real-time inspection while pouring is still underway, the feed-forward layer is intentionally compact. Weighted temporal pooling then produces the quality feature for a single pouring window. The weights are learned from pressure stability, temperature contrast, and zone-transition state, so short but important abnormal events are not smoothed away. The quality-scoring head outputs probabilities for qualified pouring, cold-joint risk, segregation risk, inadequate vibration, and unstable pumping.

Weighted temporal pooling is used to get the quality characteristic of a single pouring window. Weights do not smooth out brief but major aberrant occurrences since they are learnt based on pressure stability, temperature contrast, and zone-transition state.

$$q_w = \sum_t \gamma_t u_t \quad \text{Eq.(7)}$$

Quality Score and Defect Warning Output

A multi-class probability vector is produced by the Quality-Scoring Head. Qualified pouring, cold-joint danger, segregation risk, inadequate vibration, and unstable pumping are among the classes. The primary class will be chosen based on the greatest risk score, even if a single window may have several warning reasons.

$$y_w = \mathrm{softmax}\left(W_q q_w + b_q\right) \quad \text{Eq.(8)}$$

Temporal consistency and classification loss are included in the training objective. In order to avoid abrupt changes in nearby windows from the same pouring region unless the process has truly altered, a consistency term is included. This section often penalises abrupt feature changes between neighbouring windows, although it nevertheless permits a sharp rise in the alert when thermal boundary evidence and pressure interruption happen at the same time.

The estimated defect probability multiplied by the current zone's construction consequence weight yields the alert score. Compared to a low-risk finishing location, a column base, beam-column connection, or lengthy wall section has a larger consequence weight.

$$R_w = \sigma\left(\beta_1 y_w^{defect} + \beta_2 z_w + \beta_3 \tau_w\right) \quad \text{Eq.(9)}$$

Results and Discussion

Dataset and Pouring-State Distribution

There are 9,800 pouring windows across 36 structural sections in the simulated dataset. Wall sections account for 31.6% of windows, slabs for 28.4%, columns for 22.0%, and beams for 18.0%. The dataset is intentionally imbalanced because wall and slab pouring often lasts longer and contains more continuous video frames. Figure 3 shows dataset composition and pouring-state distribution. Figure 3(a) indicates that wall and slab windows form the two largest groups, so the training data contain more long-duration pouring scenes than short member segments; columns and beams remain smaller but necessary groups because their geometry and pouring rhythm create different thermal-pressure responses. Figure 3(b) separates the pumping process into stable, oscillation, pause, and resistance states, and the stable category reaches 56%, which means the model must avoid being biased toward normal delivery while still detecting short abnormal pressure windows. Figure 3(c) compares thermal-response groups across member types, showing that uniform responses dominate walls and slabs, whereas delayed and boundary responses become more visible in columns and beams; this pattern explains why thermal video alone may miss risk when surface temperature changes lag behind pump-pressure disturbance. Figure 3(d) reports quality-risk labels, where normal windows account for 0.62 and abnormal categories such as segregation, low vibration, cold joint, and review cases occupy smaller shares, so the evaluation must consider minority-risk recognition rather than only overall accuracy. The four views show that the evaluation is not simple visual classification, because the model must jointly interpret member shape, pressure behavior, temperature evolution, fault-label distribution, and construction consequence [21].

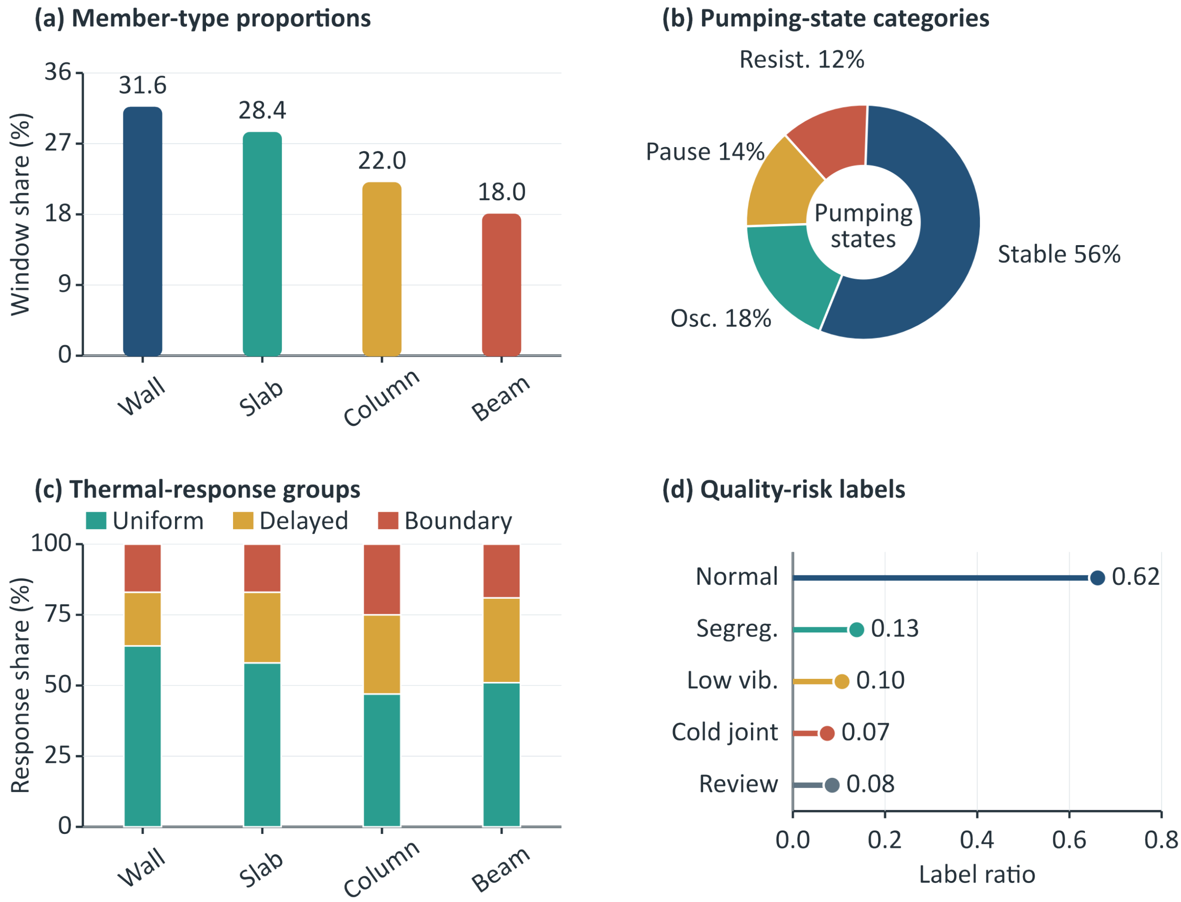

Figure 3. Dataset composition and pouring-state distribution. (a) Member-type proportions. (b) Pumping-state categories. (c) Thermal-response groups. (d) Quality-risk labels.

Their performance is compared in Table 1. Because its numerical values must be read in conjunction with the complexity of the simulated records, the table is located in the dataset section. Due to its inability to monitor the thermal boundary following a pause, a pressure-only model offers strong unstable-pumping detection but is unable to detect many cold-joint situations. A thermal-only model detects surface irregularities, but because of variations in the surrounding temperature, it generates false alerts. Although the sequence outcome is improved by LSTM and temporal convolution, its cold-joint recall is still lower than that of the suggested technique due to the weakening of various long-range pause effects during hidden-state compression. The S4D-Transformer has achieved 41 ms latency, 90.7% cold-joint recall, and 92.8% total accuracy. The statistics demonstrate that the model is still quite compact and appropriate for analysing new flaws very instantly. Among these, fewer false alarms are a practical indicator rather than a supporting statistic. The site engineers will disregard the future alarms if the warning system consistently reports typical temperature variations as quality issues. With a lower false alarm rate of 6.9%, the output is more suited for construction management. Recall and latency should be viewed in tandem, as the table illustrates. Although the Vision Transformer baseline's visual representation is superior to that of the thermal CNN, it is less appropriate for continuous inspection in many pouring zones because to its 57 ms delay and greater false alarm rate. Although temporal convolution is quick, it lacks enough memory to mitigate the quality risk associated with pauses. The suggested model has maintained excellent defect recall while keeping latency within the predicted 120 ms review budget. The balance is necessary so that the site manager won't have to make a snap decision after the concrete has been poured when it comes time to place the concrete for the following floor. The figures demonstrate that the suggested approach also enhances the class that is challenging to achieve in the field. Numerous variants have been categorised because normal pouring windows are common and visually stable. There are fewer cold-joint windows, their boundary signals are weaker, and they can be used in conjunction with formwork shadow or pump adjustment. Therefore, enhancing a large number of easy-normal samples is not as valuable from an engineering standpoint as boosting the recall of this class.

Table 1. Quantitative comparison of concrete pouring quality assessment models.

Model	Accuracy (%)	Cold-Joint Recall (%)	False Alarm (%)	Latency (ms)
Pressure Only	78.9	66.8	12.7	12
Thermal CNN	82.4	73.5	11.8	29
LSTM Fusion	86.1	81.2	9.6	38
Temporal Conv.	87.6	83.7	8.9	34
ViT Fusion	89.4	86.3	8.1	57
S4D-Transformer	92.8	90.7	6.9	41

Class overlap is a second dataset observation. Unstable pumping and cold-joint risk are connected yet distinct. The thermal field can be constant if the subsequent layer is positioned within the permissible range, and certain pumping breaks are scheduled for hose movement. On the other hand, when concrete spreading is delayed and the reinforcing density is large, a surface thermal boundary will emerge shortly after. As a result, rather than having a single value, the labels are constructed from many circumstances. The complimentary signals can minimise ambiguity, as demonstrated by sensor-fusion studies [22], and the simulated records use this strategy by maintaining the separation of the heat, pressure, vibration, and zone descriptors until model fusion. Samples at beam-column junctions and long-wall transitions are the most challenging. In certain situations, there may be a brief high-resistance segment in the pressure sequence, although the surface temperature varies slowly. The model lowers unnecessary alert production by learning to regard this combination as a possible review subject rather than a verified problem.

Controlled disruptions are also included in the dataset. There are 6.2% outside cooling disturbance windows, 10.4% formwork-shadow windows, and 7.8% camera jitter windows in thermal frames. Short missing parts, cylinder-switching pulses, and simulated variations in pipe resistance are examples of pump-pressure recordings. To stop the model trained just on clean data from doing abnormally well, add a disturbance. Evaluation should take non-stationary signal behaviour and missingness into account, as demonstrated by industrial sequence analysis [23]. In this experiment, the disturbance labels are utilised to verify resilience rather than as the final quality labels. Distinguish between a building flaw and the sensor disturbance. The two are not the same; an engineering warning is necessary for a true pouring problem, but a low-quality input should lead to a decrease in review confidence. The suggested model's ability to preserve this distinction in the face of noise is examined in the following graphics. The fact that pressure and heat signals fail in distinct ways is another justification for adding a disturbance. A darkened region will cause a thermal camera's contrast to decrease, but the delivery cycle may still be seen in the pressure curve. The thermal field can still show if concrete installation is going as planned even if there is a short in the pressure sensor during pump changeover. As a result, the dataset evaluates complementing recovery as opposed to only mean noise tolerance. Automatic alerts are not employed since the design is particularly appropriate for the issue of unpredictable input-signal quality at the site. Additionally, a disturbance set will prevent the model from learning a shortcut. The model would not be able to identify pump continuity and member position if every cold spot was designated as defective. The model would classify the typical cylinder changeover as a quality issue if every pressure peak was identified as a malfunction. As a result, the model must understand the connection between the physical process and observable reaction since the simulated samples contain counterexamples, such as cold but continuous surfaces and high-pressure but well-compacted placement windows.

Quality Prediction and Defect Recognition

The first predicted results are shown in Figure 4. Figure 4(a) shows the general quality-classification accuracy; the proposed model is better than pressure-only, thermal-only, LSTM, temporal convolution and Vision

Transformer baselines. Figure 4(b) is the cold-joint recall rate, and a missing placement discontinuity may not be discovered until the final check. Figure 4(c) shows the reduction of false alarms, and cross-modal fusion avoids raising alarms for transient pressure or temperature fluctuations. Based on the above, the model has increased both the average prediction accuracy and the usefulness of site reviews, which is consistent with other construction analytics studies that consider both operating cost and defect sensitivity [24].

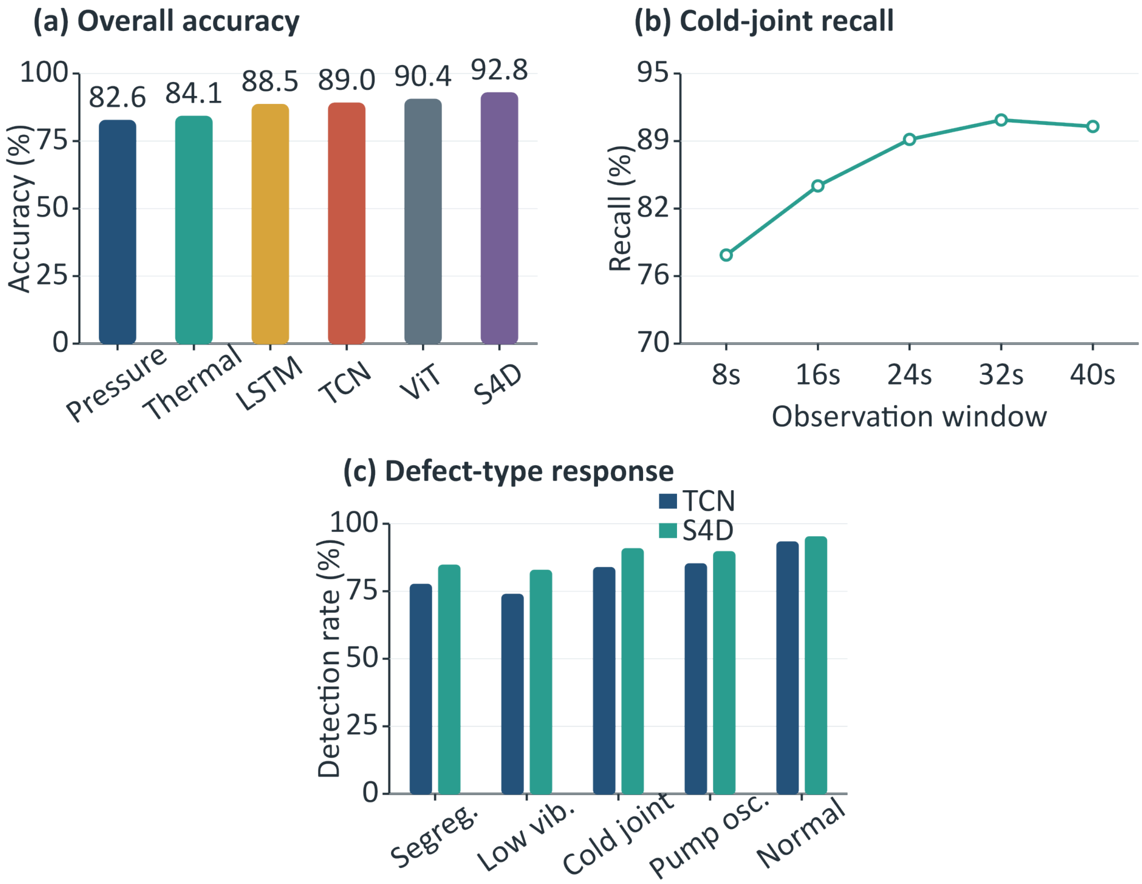

Figure 4. Model prediction performance. (a) Overall accuracy. (b) Cold-joint recall. (c) Defect-type response.

The combined thermal-pressure evidence is as follows: Figure 5. Figure 5(a) connects the thermal path with pressure-pulse timing to show whether a surface-temperature boundary forms after a pumping pause. Figure 5(b) shows the evolution of the quality score; it increases only when multiple parts of the process evidence agree. Therefore, it cannot be concluded that a single-source anomaly is, in fact, a quality defect. Temporary pressure oscillation may be process-related; a pause accompanied by local cooling, poor vibration response and a zone transition are more indicative of a cold-joint risk. Therefore, Figure 5 supports traceable engineering review instead of opaque classification and follows the extended cross-modal sequence-fusion argument that local attention and long-memory cues should be combined for complex temporal evidence [25].

https://doi.org/10.64972/dea.2026.v5i2.3958d:101-113

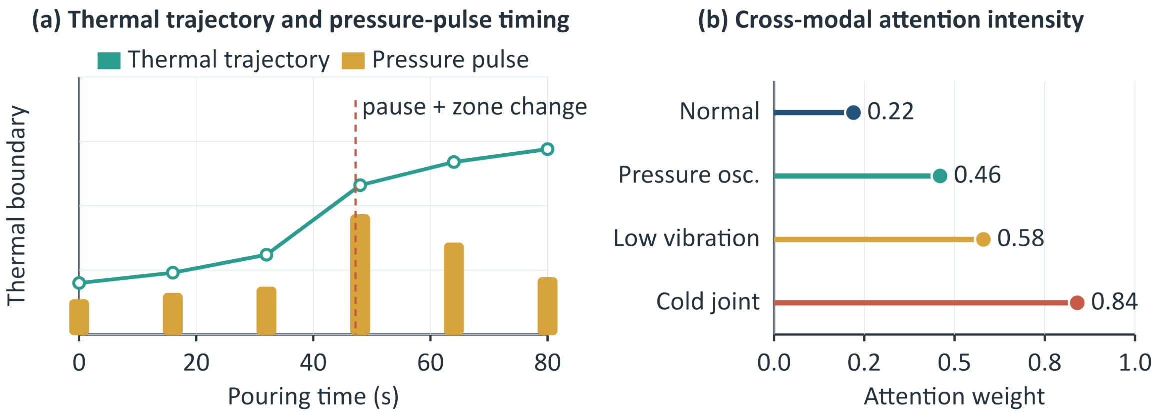

Figure 5. Thermal-pressure coupling interpretation. (a) Thermal trajectory and pressure-pulse timing. (b) Cross-modal attention intensity.

In the typical pouring area, the Quality Score is steady, but it increases dramatically when there are several risk flags. During a brief hose displacement in a single simulated wall section, the temperature field is constant and the pump pressure rises from 5.8 MPa to 8.3 MPa. The condition is categorised by the model as pressure oscillation rather than cold-joint danger. The warning score rises to 0.84 in another section after a 96-second wait, a local chilly boundary, and a poor vibration response. As a result, neither a signal abnormality nor a quality flaw exists in this instance. A defect alert necessitates evidence that the material condition or placement continuity has been compromised, whereas an anomaly may be transient and process-related. Studies on construction inspections [26] demonstrate that the automated evaluation utilised in engineering review requires interpretable evidence. Consequently, rather than just returning a final class, the model output maintains the pressure event, temperature boundary, zone identification, and warning score together.

There is a use for the confusion pattern as well. The majority of the mistakes are caused by a combination of low vibration and high segregation risk, which both produce uneven surfaces. They may be distinguished by the pressure sequence: inadequate vibration often results in a steady delivery but a poor vibration response, and pressure fluctuation and uneven thermal streaks increase the danger of segregation. Due to thermal visibility restrictions brought on by formwork and reinforcement shadows, the suggested model still incorrectly classifies some edge situations close to the column bases. These examples demonstrate when the review interface needs to ask for manual approval; they do not invalidate the model. The simulated findings support the idea that autonomous systems should incorporate prediction and inspection procedures, as suggested by research on real-time construction monitoring [27]. As a result, rather of being a definitive technical judgement, a warning is given as a review item with signal evidence. A false positive and a false negative have distinct site costs, therefore this interpretation is necessary for the concrete job. A false negative will permit a problem to persist in the product till later, whereas a false positive will halt the pour and increase labour coordination costs. As a result, the model will be assessed concurrently on recall, false alarm, and review load. When the camera view is partially blocked and the pump hose is moved quickly, the short column-base window has the greatest remaining error rate in the simulated records. Future site usage should incorporate camera location planning, structured inspection notes, and thermal-pressure analysis based on the samples. The output class should be linked to a likely field activity based on the analysis. Placement discontinuity and layer connection concerns are indicated by a cold-joint warning; compaction review is indicated by an insufficient-vibration warning; mixture non-uniformity is shown by a segregation warning; and pump-line resistance or delivery rhythm problems are indicated by an unstable-pumping warning. Thus, the issue of site response and model prediction has been resolved.

Ablation, Robustness, and Deployment Analysis

Ablation and robustness analysis are shown in Figure 6. When the S4D block is removed, as shown in Figure 6(a), the model loses memory of earlier pumping pauses and cold-joint recall drops. Long wall sections are most affected because the defect signal is spread across several windows. When Transformer fusion is removed in Figure 6(b), false alarms increase because thermal and pressure cues are no longer matched through cross-modal attention. Figure 6(c) evaluates thermal-noise robustness, and Figure 6(d) evaluates pressure-missingness robustness. The revised interpretation emphasizes that absent pressure data should reduce warning confidence rather than be hidden by overconfident interpolation [28].

Model performance and deployment interpretation are linked in Figure 7. Figure 7(a) reports inference latency under single-zone, four-zone, and eight-zone processing. The proposed model remains within the simulated 120 ms review budget, with 41 ms per window for single-zone operation and 86 ms for eight-zone batching. Figure 7(b) reports warning-priority distribution, including normal review, observation, engineering review, and emergency warning. This distribution is necessary because a practical site system should not trigger an alarm for every minor disturbance. The warning queue is sorted by quality risk and construction consequence weight, so a beam-column joint with moderate defect probability may be prioritized above a low-risk slab finishing area [29].

Figure 6. Ablation and robustness analysis. (a) S4D block removal. (b) Transformer fusion removal. (c) Thermal-noise robustness. (d) Pressure-missingness robustness.

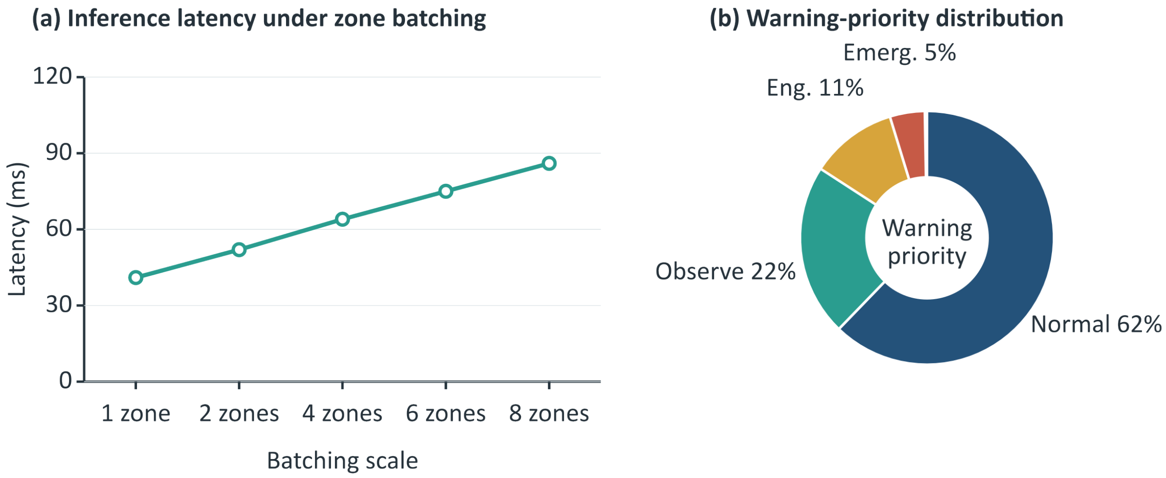

Figure 7. Field-oriented deployment analysis. (a) Inference latency under zone batching. (b) Warning-priority distribution.

https://doi.org/10.64972/dea.2026.v5i2.3958d:101-113

Traceable deployment is necessary. A site engineer should be able to identify the pouring zone, pressure segment, and thermal region that caused a warning. Explainable decision-support research [30] suggests that users trust automated systems more when evidence and uncertainty are visible. The proposed model therefore outputs a risk score, defect type, confidence level, thermal image region, pressure interval, and construction-zone tag. An unstable-pumping alert may require checking pipe resistance, pump rhythm, or mixture consistency, while a cold-joint warning near a long wall may require immediate review of placement continuity. Anomaly-detection surveys [31] also show that tail risk, false alarms, and review burden should be reported rather than relying only on average performance.

The goal of the S4D-Transformer model is to transform scattered site signals into a comprehensible construction-quality record. Thermal video provides local surface evidence, pump pressure provides delivery-process evidence, vibration response provides compaction evidence, and zone metadata provides engineering-consequence evidence. Human-centered AI research shows that decision-support systems should complement expert judgment rather than replace it [32]. The revised deployment view therefore includes a review queue and signal explanations. Each warning report contains a thermal image region, pressure segment, zone identifier, risk score, confidence level, and recommended review type. Engineering inspection remains necessary for final quality judgment [33], sample data should be retained to support later verification [34], and raw thermal-pressure records should be preserved for dispute review or quality tracing [35].

Conclusion

This study presents an S4D-Transformer model for concrete pouring quality assessment using thermal video and pump-pressure data. The method constructs synchronized pouring windows, extracts thermal texture and pressure-pulse features, models long-range pouring states with S4D, and uses Transformer attention to fuse cross-modal evidence. In the simulated dataset, the model outperforms pressure-only, thermal-only, LSTM, temporal convolution, and Vision Transformer baselines, increasing overall classification accuracy to 92.8% and cold-joint recall to 90.7%. The main contribution is to treat pouring inspection as process-level sequence evaluation rather than isolated image or pressure classification.

First, it is possible to think of concrete-pouring inspection as a sequence evaluation problem at the process level. Vibration response, pump-pressure curves, thermal pictures, and structural-zone information are not separate sources of proof. They depict several stages of the same building process. In order to differentiate between defect-related interruption and intended pumping adjustment, as well as between relevant cold-boundary development and ordinary thermal fluctuation, the proposed model combines long-range state memory with local cross-modal attention.

Several limitations remain. Because the dataset is simulated, it cannot fully represent weather variation, camera occlusion, mix-design diversity, formwork geometry, pump equipment differences, vibration practice, manual operating habits, and project-specific acceptance criteria. A practical implementation should include calibrated thermal cameras, validated pump-controller records, vibration sensor logs, field inspection labels, post-hardening inspection results, and project-specific quality thresholds. Future site application should also preserve the original thermal video and pressure history so that warnings can be audited during dispute review or quality tracing. These additions are needed before the model can move from controlled simulation to real construction-quality management.

Author Contributions

Sultan Al-Dhaheri contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Abdullah Al-Zaabi contributes to validation, analysis, investigation, data collection, draft preparation. Juma Al-Sharqi contributes to software, validation, analysis, investigation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

References

[1] Namatēvs, I., Jēkabsons, A., Rjabcevs, A., & Noviks, G. (2025). ConMonity: An IoT-Enabled LoRa/LTE-M Platform for Multimodal, Real-Time Monitoring of Concrete Curing in Construction Environments. Sensors, 26(1), 14. https://doi.org/10.3390/s26010014

[2] Pozzer, G., Giordano, A., & Tubaldi, E. (2024). Enhancing concrete defect segmentation using multimodal data and Siamese Neural Networks. Automation in Construction, 167, 105594. https://doi.org/10.1016/j.autcon.2024.105594

[3] Deng, F., Xu, L., Chen, Y., & Zhang, J. (2024). Determination of Lubrication Layer Thickness and Its Effect on Concrete Pumping Pressure. Materials, 17(20), 5136. https://doi.org/10.3390/ma17205136

[4] Kang, M., & Kang, D. (2024). Transformer-based multivariate time series anomaly detection using inter-variable attention mechanism. Knowledge-Based Systems, 290, 111507. https://doi.org/10.1016/j.knosys.2024.111507

[5] Wen, M., Cheng, J. C. P., & Chen, K. (2024). Automated construction safety reporting system integrating deep learning-based real-time advanced detection and visual question answering. Advances in Engineering Software, 196, 103779. https://doi.org/10.1016/j.advengsoft.2024.103779

[6] Chen, H., Wang, X., Zhang, Y., & Li, Q. (2024). Thermal property evolution and prediction model of early-age low-heat cement concrete under different curing temperatures. Journal of Building Engineering, 82, 108020. https://doi.org/10.1016/j.jobe.2023.108020

[7] Mikhalev, M., Secrieru, E., & Mechtcherine, V. (2024). Practical Insights and Advances in Concrete Pumping. RILEM Technical Letters, 9, 198. https://doi.org/10.21809/rilemtechlett.2024.198

[8] Xiao, B., Chen, J., Li, H., & Cheng, J. C. P. (2024). Automated daily report generation from construction videos using ChatGPT and computer vision. Automation in Construction, 169, 105874. https://doi.org/10.1016/j.autcon.2024.105874

[9] Ahamed, M. A., & Cheng, Q. (2024). TimeMachine: A Time Series is Worth 4 Mambas for Long-Term Forecasting. Frontiers in Artificial Intelligence and Applications, 392, 677-684. https://doi.org/10.3233/FAIA240677

[10] Wang, Z., Ma, C., Zhou, Y., & Li, X. (2025). Is Mamba effective for time series forecasting? Neurocomputing, 619, 129178. https://doi.org/10.1016/j.neucom.2024.129178

[11] Zhu, Y., & Wang, Y. (2024). Experimental modeling and quantitative evaluation of mitigating cracks in early-age mass concrete by regulating heat transfer. Journal of Building Engineering, 98, 110641. https://doi.org/10.1016/j.jobe.2024.110641

[12] Guo, J., Liu, Y., Zhang, H., & Chen, X. (2024). Automatic assessment of concrete cracks in low-light, overexposed, and blurred images restored using a generative AI approach. Automation in Construction, 168, 105787. https://doi.org/10.1016/j.autcon.2024.105787

[13] Kang, D., Kim, S., & Kim, J. (2024). Automatic detection and classification process for concrete defects in deteriorating buildings based on image data. Journal of Asian Architecture and Building Engineering. https://doi.org/10.1080/13467581.2024.2373832

[14] Chen, X., Li, Y., Wang, J., & Zhang, T. (2025). Experimental Study on Pulse Elimination of Concrete Pumping. Journal of Construction Engineering and Management, 151(3), 04025001. https://doi.org/10.1061/JCEMD4.COENG-15462

[15] Yuan, B., Xu, J., & Liu, K. (2024). Insights and challenges of predicting concrete pumpability: A state-of-art review. Journal of Building Engineering, 97, 110265. https://doi.org/10.1016/j.jobe.2024.110265

[16] Wu, X. (2025). Intelligent Construction Safety Warning and Risk Assessment Model Based on Multi-Source Sensor Data Fusion. 2025 2nd International Conference on Telecommunications and Power Electronics, 43-48. https://doi.org/10.1109/TELEPE66277.2025.00043

[17] Kim, S., Park, H., & Choo, J. (2024). When Model Meets New Normals: Test-Time Adaptation for Unsupervised Time-Series Anomaly Detection. Proceedings of the AAAI Conference on Artificial Intelligence, 38(12), 13213-13221. https://doi.org/10.1609/aaai.v38i12.29210
[18] Wu, Y., Chen, J., Wang, H., & Liu, S. (2026). DTMamba: Dual Twin Mamba for Time Series Forecasting. Tsinghua Science and Technology, 31(1), 143-156. https://doi.org/10.26599/TST.2024.9010143
[19] Feng, Y., Zhang, L., Chen, X., & Wang, M. (2025). DecMamba: Mamba Utilizing Series Decomposition for Multivariate Time Series Forecasting. Computers, Materials & Continua, 82(2), 2219-2238. https://doi.org/10.32604/cmc.2024.058374
[20] Pessoa, L., Santos, R., Silva, M., & Costa, A. (2025). Mamba time series forecasting with uncertainty quantification. Machine Learning: Science and Technology, 6(2), 025018. https://doi.org/10.1088/2632-2153/adec3b
[21] Senthilnathan, K., & Raphael, B. (2024). Buildability assessment of 3D printed concrete elements through computer vision. Proceedings of the International Symposium on Automation and Robotics in Construction, 41, 1015-1022. https://doi.org/10.22260/ISARC2024/0129
[22] Kim, D., Lee, S., Park, J., & Choi, H. (2025). Point Cloud-Based Defect Inspection with Multisensor Fusion and Deep Learning for Advancing Building Construction Quality. Journal of Computing in Civil Engineering, 39(2), 04025002. https://doi.org/10.1061/JCCEE5.CPENG-6060
[23] Xiong, Y., Zhang, J., Liu, H., & Chen, W. (2024). SiET: Spatial information enhanced transformer for multivariate time series anomaly detection. Knowledge-Based Systems, 294, 111928. https://doi.org/10.1016/j.knosys.2024.111928
[24] Wang, Z., & El-Gohary, N. M. (2024). Few-shot object detection and attribute recognition from construction site images for improved field compliance. Automation in Construction, 166, 105539. https://doi.org/10.1016/j.autcon.2024.105539
[25] Xia, Y., Zhang, H., Liu, J., & Wang, C. (2024). MFAM-AD: an anomaly detection model for multivariate time series using attention mechanism to fuse multi-scale features. PeerJ Computer Science, 10, e2201. https://doi.org/10.7717/peerj-cs.2201
[26] Fu, X., Wang, L., Chen, S., & Zhao, Y. (2024). Multivariate time series anomaly detection via separation, decomposition, and dual transformer-based autoencoder. Applied Soft Computing, 157, 111671. https://doi.org/10.1016/j.asoc.2024.111671
[27] Guo, Z., & Jiao, D. (2024). Rheology as a versatile tool in concrete technology: a mini-review. Advanced Manufacturing, 12, 356-370. https://doi.org/10.55092/am20240005
[28] Liang, Y., Chen, Z., Wang, T., & Liu, X. (2024). Anomaly Detection in Time Series Data. International Journal of Prognostics and Health Management, 15(3), 1-17. https://doi.org/10.36001/ijphm.2024.v15i3.3853
[29] Mittal, S. (2024). A comprehensive survey of deep learning-based lightweight object detection models for edge devices. Artificial Intelligence Review, 57, 311. https://doi.org/10.1007/s10462-024-10877-1
[30] Coussement, K., Phan, M., Benoit, D. F., & De Caigny, A. (2024). Explainable AI for enhanced decision-making. Decision Support Systems, 181, 114276. https://doi.org/10.1016/j.dss.2024.114276
[31] Zamanzadeh Darban, Z., Webb, G. I., Pan, S., Aggarwal, C. C., & Salehi, M. (2024). Deep Learning for Time Series Anomaly Detection: A Survey. ACM Computing Surveys, 57(1), Article 9. https://doi.org/10.1145/3691338
[32] Zhou, Y. (2025). Multi-dimensional model and interactive simulation of intelligent construction based on digital twins. Scientific Reports, 15, 28941. https://doi.org/10.1038/s41598-025-17100-3
[33] Baniecki, H., & Biecek, P. (2024). Adversarial attacks and defenses in explainable artificial intelligence: A survey. Information Fusion, 107, 102303. https://doi.org/10.1016/j.inffus.2024.102303
[34] Nam, J., Kim, H., Lee, J., & Kang, U. (2024). Breaking the Time-Frequency Granularity Discrepancy in Time-Series Anomaly Detection. Proceedings of the ACM Web Conference 2024, 2458-2469. https://doi.org/10.1145/3589334.3645556
[35] Li, Y., Zhang, X., Wang, S., & Chen, L. (2024). An explanation framework and method for AI-based text emotion analysis and visualisation. Decision Support Systems, 177, 114121. https://doi.org/10.1016/j.dss.2023.114121