Scholar Publishing Group
International Journal of Neural Network
https://doi.org/10.38007/NN.2026.050113
ISSN 2790-2692 Vol. 5, Issue 1: 119-129

Design of Infrared Thermal Imaging Fault Detection and Early Warning System for Data Center Equipment

Xinyi Sun

NYU Tandon School of Engineering, New York University, Brooklyn, 11201, New York, US

Keywords: Data Centre, Infrared Thermal Imaging, Equipment Fault Detection, Early Warning, Thermal Anomaly Identification, Edge Intelligence

Abstract: To address the hidden heating hazards in high-density server racks at power distribution units, UPS terminals and busbar connections, this paper introduces an infrared thermal imaging-based equipment fault detection and multi-level early warning system. The system has integrated long-wave infrared cameras, visible-light assisted asset mapping, rack power, ambient temperature and humidity, local airflow, and maintenance feedback to form a closed-loop architecture of "thermal image acquisition-radiation calibration-target area location-temperature rise anomaly identification-graded warning linkage". Methodologically, this paper builds a radiance-corrected temperature inversion model, a load-corrected temperature-rise residual model and a risk integral model with duration filtering. Literature indicators are redrawn according to task type, and classification accuracy, detection mAP and segmentation Jaccard are no longer shown in the same chart. A simulated rack-aisle case further confirms that transient hot spots are filtered, and sustained PDU terminal loosening can be upgraded from attention to an alarm based on the temperature-rise and duration standards.

1 Introduction

A typical Data Centre has server racks, PDUs, UPSs, battery banks, rack cabinets, air-conditioning terminals and an integrated cabling system. They have a high-load density, strong spatial coupling, and run continuously in the cloud. With the development of AI training, cloud computing and edge services, the power density per rack has been continuously rising. An increase in the local connection resistance, irregular airflow organisation, loose terminals, fan attenuation and blocked heat dissipation channels can all lead to the development of hot spots. Zhao and others have systematically studied local hot-spot elimination technologies for data centers, and now thermal management problems are no longer limited to energy-efficient air conditioning; they have become all-encompassing safety issues that affect equipment reliability, room layout and airflow organisation [1]. In a high-density deployment scenario, if a single-point temperature increase cannot be identified in time, it will accelerate insulation ageing, increase contact resistance, reduce server frequency, trigger UPS bypass switching, or cause a rack-level service interruption.

Traditionally, the inspection of computer rooms has only used manual handheld thermal imaging devices, a small number of temperature- and humidity-sensing probes, and sensors embedded in the equipment. The mode can find obvious hot spots, but it has the following defects: a long inspection cycle; insufficient coverage of measurement points; discontinuous image evidence; and subjective differences in judgment. Chen and others have used simulation data and machine learning to study the prediction of rack hot spot temperature; airflow speed, server power and supply air temperature are correlated to form the rack's entrance thermal environment, and local hot spots can be estimated by data-driven models [2]. The above methods generally focus more on predicting the thermal environment of the rack and do not provide sufficient support for infrared image-level fault identification of terminals, busbars, PDU sockets and internal UPS connections. Wiysobunri et al. have integrated infrared thermal imaging and CNN for server status monitoring and fault diagnosis, and provided direct evidence of intelligent thermal imaging detection of data centre equipment [3]. It can be seen that the functions of an infrared thermal imaging early warning system for data center equipment should include temperature measurement, target positioning, load correction, anomaly classification, and operation-and-maintenance closed-loop capabilities, etc.

This paper investigates how to apply infrared thermal imaging as an online fault early warning system for data center cabinets, PDUs and UPS equipment, and designs a thermal anomaly detection architecture for this purpose. The main contributions of this paper are as follows: First, a multi-level online infrared thermal imaging architecture has been proposed, and the boundaries of the sensing layer, edge inference layer, status assessment layer and alarm linkage layer have been explicitly defined; Second, temperature calibration, component-region features and load-corrected risk integral models have been built to judge absolute temperature, relative temperature rise and anomaly persistence simultaneously; Third, public literature data have been remapped according to task type, and direct comparisons of Accuracy, mAP and Jaccard across different recognition tasks have been avoided; Fourth, engineering alarm criteria have been specified, including concrete temperature-rise ranges, maximum temperature references and persistence windows; Fifth, a rack-aisle simulation case has been introduced to verify the filtering effect on transient thermal interference and the escalation capability for sustained terminal overheating.

2. Current Status Analysis of the Research Topic

Non-destructively diagnose the state of equipment with an infrared thermal imager to show pictures, positions and sensitivities to early-stage temperature changes. Tang and Jian proposed an infrared image recognition and diagnosis method for complex electrical equipment, and reported the recognition accuracy and false alarm rate of several kinds of components to provide a quantitative basis for identifying thermal defects in electrical equipment [4]. A strength of the above research is that it can directly extract the spatial morphology of the fault from thermal images and reduce dependence on human experience; however, a weakness is that most of the sample sources are concentrated in substations, switch cabinets, or general electrical equipment, which are still different from the actual conditions of internal obstruction, air duct disturbance, and dense arrangements of multiple devices in data center cabinets.

Lightweight neural networks are now used to some extent for online thermal image detection. Zhou and others have proposed a lightweight convolutional neural network for electrical equipment in substations that reduces model complexity while maintaining recognition performance, and shown that thermal image fault detection can be deployed on resource-constrained terminals [5]. Li and others have designed PEDNet based on YOLOv4-Tiny for power equipment target detection in infrared images and shown that a light-weight target detection network can be deployed at the edge [6]. Xiang and others used the improved YOLOv8 model for infrared target detection of substation equipment and further showed that multi-scale feature fusion and small target enhancement are necessary for infrared equipment detection [7]. However, the main targets in the data centre sites are not large exposed equipment, but plugs, terminals, cable connections, UPS module edges and local air vents in cabinets. The target scale is small and the background thermal texture is more complex; therefore, the power station model cannot be used directly.

Locate the defect area and determine if it is a type of fault. Siami and others used thermal image enhancement and a U-Net network for thermal defect segmentation of conveyor belt idlers, reported a high pixel-level segmentation index, and thus refined the boundaries of the thermal defect regions through semantic segmentation [8]. Resendiz-Ochoa et al. have used thermal imaging and convolutional neural networks for fault diagnosis of motors and gearboxes, achieved good results in terms of accuracy, precision, recall and F1 score, and shown that infrared thermal imaging has transferable potential for the fault diagnosis of electromechanical equipment [9]. Yao and others have built an automatic detection method for the diagnosis of cable terminal overheating and presented mAP@0.5 and detection time index, providing a performance benchmark for the design of a real-time early warning system [10]. In short, although recent studies have shown good results in infrared thermal imaging intelligent recognition for electrical and electromechanical equipment, there are three main deficiencies: First, most studies use offline datasets and lack concept drift processing under continuous operation; Second, the algorithm indicators generally focus on detection accuracy and rarely integrate with work orders, DCIM, BMS and equipment asset management; and Third, insufficient attention has been paid to temperature baseline drift caused by load changes, air conditioning strategies and equipment aging.

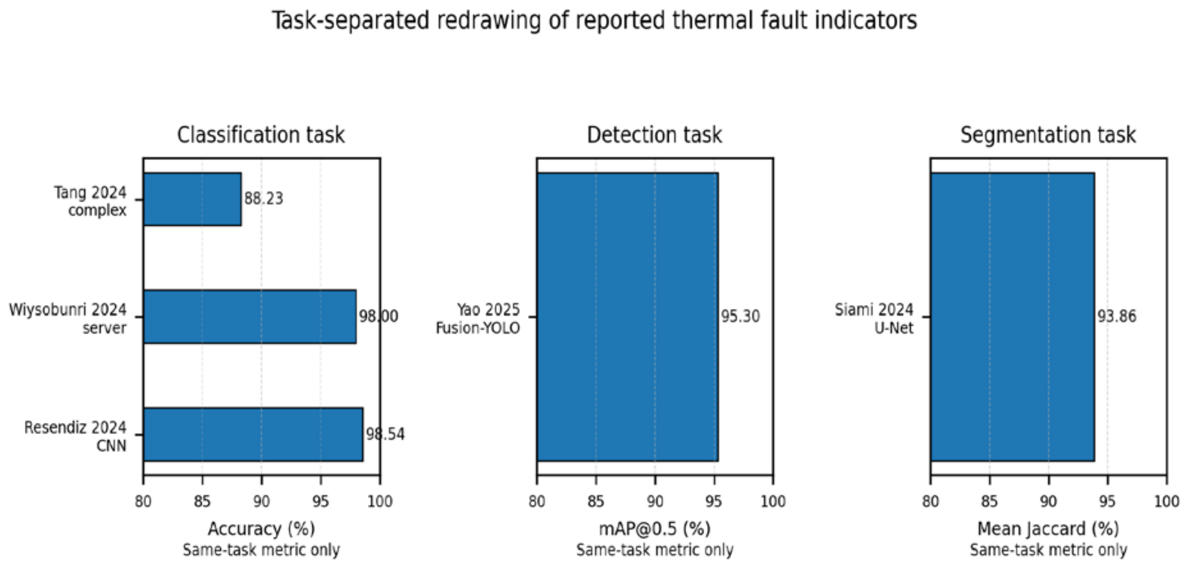

Figure 1. Task-Separated Redrawing of Publicly Available Indicators for Thermal Fault Identification.

Figure 1 is a redrawing of the quantifiable thermal fault diagnosis results published in the past three years by task type in English literature. Classification studies are shown with Accuracy; detection studies with mAP@0.5; and segmentation studies with mean Jaccard. Thus, the three indicators are only to be considered within their own scopes. This task-separated presentation avoids using heterogeneous metrics as the single ranking scale and only shows the range of evidence that can support data center thermal fault detection design. For engineering deployment, the literature still needs to be linked with real-time latency, false alarm rate, cross-rack robustness and alarm handling cost.

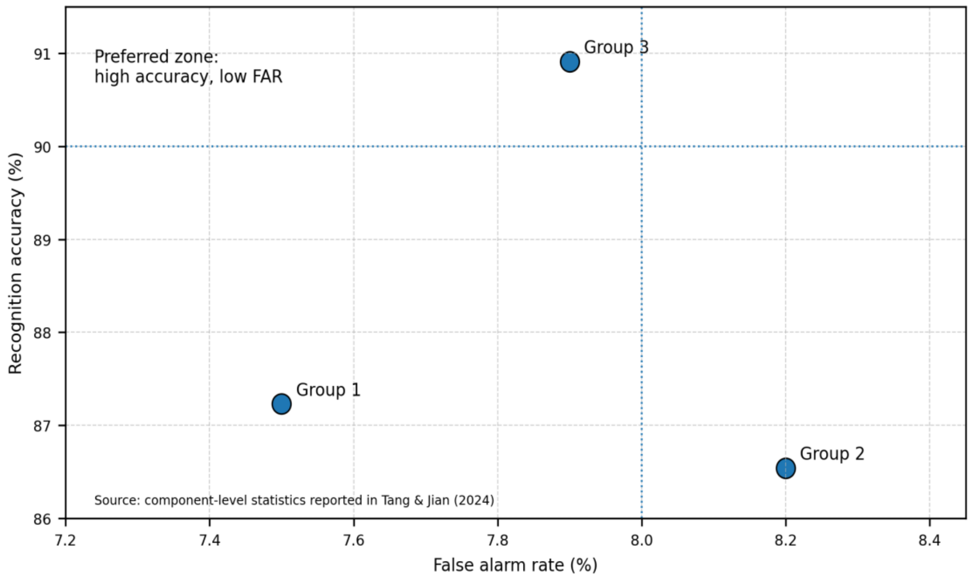

Figure 2. Same-task Redrawing of Recognition Accuracy and False Alarm Rate.

Figure 2 keeps the comparison at the same recognition task and redraws the component-level recognition accuracy and false alarm rate reported by Tang and Jian. The desired engineering direction is high recognition accuracy and a low false alarm rate; otherwise, false alarms in data centre equipment will lead to repeated dispatches and manual inspections, and overlooked overheating faults will accumulate risks along the power supply path. Therefore, the system should use a combination of model confidence, load-corrected temperature rise, duration filtering and multi-source evidence confirmation instead of a single-frame probability or a fixed temperature threshold.

Table 1. Hardware and Deployment Parameters of the Data Center Infrared Thermal Imaging Early Warning System

Layer	Component	Key specification	Role
Sensing	LWIR thermal camera	8-14 um, 640 x 512, <=50 mK NETD, 9-30 FPS	Rack, PDU and UPS thermal stream
Alignment	Visible-light camera	2-8 MP, PoE, fixed viewpoint	ROI verification and asset mapping
Edge computing	Edge GPU node	TensorRT, 8-16 GB GPU memory, SSD cache	Local inference and alarm filtering
Integration	DCIM/BMS gateway	MQTT, OPC UA, REST API	Alarm delivery and ticket linkage
Storage	Time-series and object storage	TimescaleDB/InfluxDB plus object store	Feature history and evidence archive
Baseline monitoring	PDU/UPS meter and airflow probe	Power, ambient T/H and local wind speed, 1-min aggregation	Load correction and baseline drift compensation
Alarm console	DCIM/ITSM evidence page	Asset ID, ROI image, temperature curve and risk grade	Closed-loop confirmation and label feedback

Table 1 is the system deployment information at the hardware level. Infrared cameras continuously collect the temperature field of the equipment, visible-light cameras are used to help position and map assets, edge nodes perform local inference and alarm filtering, and the gateway synchronizes the results with DCIM or BMS. The purpose of the above architecture is to build an integrated data chain for thermal images, assets, load data and alarms, and thus solve the problems of fragmented evidence and unclear responsibility in manual inspections.

3. Raise questions

Core Problems of Infrared Thermal Imaging Fault Detection for Data Centre Equipment. First, the Environment is relatively warm-tempered. The same PDU contact will have a different temperature under different loads of the rack, supply-air temperature and airflow. Only a fixed-temperature upper bound is used to identify high-load conditions as failures incorrectly. Second, the target areas in infrared images are small and frequently obstructed. Data centre cables, cable ties, plugs, metal edges of equipment create a complex background that causes a performance drop for target detection models during rack migration. Thirdly, the equipment failures often occur in stages. Poor contact, loose terminals and fan attenuation generally show a slow rise in temperature rather than a sudden drop; therefore, single-frame classification may fail to detect prolonged abnormal conditions. Finally, the engineering alerts must be feasible. Maintenance staff need to know the location of the anomaly, the temperature difference, the risk level, recommended actions and supporting images; model probability values alone are not sufficient.

Therefore, this paper defines the problem as follows: Given a continuous infrared image sequence, equipment asset mapping, environmental parameters and load information, identify areas of abnormal temperature rise in key components of cabinets, PDUs and UPSs, calculate interpretable risk scores, and trigger multi-level alarms when persistence and reliability constraints are met. The purpose of this problem is not to maximise the accuracy of offline classification, but rather to achieve early fault detection under conditions of controllable false alarms, good real-time performance, and executable closed-loop operation and maintenance.

4. Problem-solving and system design strategies

4.1 Overall Architecture and Data Flow

The four levels of the system are: the perception layer, the edge computing layer, the status assessment layer and the alarm linkage layer. The deployment unit is a rack aisle, and a single camera does not constitute one. Each fixed infrared camera covers the rear area of two to four adjacent racks, and a visible-light view is used to bind the thermal ROI with the PDU socket, UPS terminal, busbar connection and cabinet asset code. The edge node performs radiation correction, ROI registration, local model inference and temporary storage in the computer room, and the status assessment layer fuses rack power, ambient temperature, humidity and airflow to calculate the load-corrected temperature rise. The alarm linkage layer will output the risk level, evidence image, temperature curve and recommended handling action to DCIM, BMS or ITSM, and thus the system output can be turned into an executable maintenance task.

Table 2 is the conversion path of raw thermal images to work-order alarms. Calibration enables temperature comparison; Registration links thermal deviations to particular assets; Feature extraction and model inference support risk assessment; Risk integration uses duration and spatial consistency to suppress transient hot spots; and alarm output includes image evidence, temperature curves and asset locations. After deployment, add quality control records to address problems such as camera drift, new racks, changed view angles and confirmed false alarms.

Table 2. Infrared Thermal Imaging Data Processing and Early Warning Software Flowchart

Stage	Input	Operation	Output	Quality control
Calibration	Raw thermal frame	Emissivity and ambient correction	Temperature matrix	Blackbody check and drift log
Registration	Thermal and RGB frames	ROI mapping and homography update	Asset-level ROI	Manual review for new racks
Feature extraction	ROI temperature matrix	Max, mean, p95, gradient and hot area	Feature vector	Missing-frame interpolation
Inference	Feature vector and image crop	CNN detection plus residual scoring	Fault probability	Confidence threshold and temporal voting
Risk integration	Fault probability, temperature rise and duration	Temporal voting and risk integral	Risk grade and persistence evidence	Duration and spatial consistency check
Warning	Risk grade, Delta T threshold and duration evidence	Multi-level rule engine	Alarm event and evidence pack	Ticket feedback and label update

4.2 Temperature Calibration and Regional Feature Modeling

Infrared cameras measure radiance, not direct temperature. Let $i,j$the observed radiance of the i-th pixel be L_{ij}, the target emissivity be ϵ_{ij}, the air transmittance be τ, the ambient reflectance be L_{amb}, and the camera calibration constants be $K_1$and K_2. The radiance after environmental and emissivity correction is L_{ij}^c, and the pixel temperature inversion can be expressed as:

$$L_{ij}^c=\frac{L_{ij}\text{-}(1\text{-}\tau)L_{amb}}{\epsilon_{ij}\tau},\quad T_{ij}=\frac{K_2}{\ln\left(\frac{K_1}{L_{ij}^c}+1\right)}\text{-}273.15 \tag{1}$$

In equation (1), $T_{ij}$is the Celsius temperature and $L_{ij}^c$is the corrected radiance. This formula is used to convert thermal images from different times and camera perspectives into comparable temperature matrices. In data center sites, the emissivity of the PDU casing, copper busbars, insulation materials, and cable surfaces differs. If the emissivity differences are ignored, the model may easily misjudge material differences as abnormal temperature rise.

After obtaining the temperature matrix, the system $r$constructs a feature vector for each device component region. Let be the region's highest temperature, $T_{r,t}^{max}T_{r,t}^{mean}$be the average temperature, $T_{r,t}^{p95}$be the 95th percentile temperature, $\Delta T_{r,t}$be the load-corrected temperature rise, $G_{r,t}$be the temperature gradient intensity, $A_{r,t}^{\text{hot}}$ and be the hotspot area percentage. Then the region characteristics are:

$$\mathrm{x}_{r,t}=\left[T_{r,t}^{max},T_{r,t}^{mean},T_{r,t}^{p95},\Delta T_{r,t},G_{r,t},A_{r,t}^{\text{hot}}\right]^{\mathrm{T}} \tag{2}$$

The features in Equation (2) are both absolute temperature and spatial morphology and hot spot area, so the highest temperature alone may be misleading due to accidental errors. For example, a single reflective point will show a local high-temperature reading, but its hot spot area and gradient continuity are low; actual poor contact generally presents as a localised sustained high temperature, a stable hot spot boundary, and an increased temperature difference in the adjacent area.

4.3 Load-corrected temperature rise and failure probability estimation

Data center equipment temperature is significantly affected by load and airflow conditions. Let be the ambient temperature of the server room, $T_{room,t}P_{r,t}$be the component-related load power, $v_{air,t}$be the local wind speed, $\alpha_0,\alpha_1,\alpha_2,\alpha_3$and be the baseline parameters fitted under historical healthy conditions. Then, the load-corrected temperature rise is defined as:

$$\Delta T_{r,t}=T_{r,t}^{max}-\left(\alpha_0+\alpha_1 T_{room,t}+\alpha_2 P_{r,t}+\alpha_3 v_{air,t}\right) \tag{3}$$

Equation (3) separates "normal heating due to a high load" from "abnormal heating relative to the baseline". If the cabinet power increases but the change in temperature falls within the range of historical deviations, a high-level alarm will not be immediately triggered; however, if the power change is relatively small and the maximum temperature exceeds the baseline significantly, it may be due to problems such as increased contact resistance, blocked heat dissipation paths or aging of local components.

Failure probability estimation employs a fusion of lightweight convolutional features and regional statistical features. Let be the temporal hidden state of the previous time step, $h_{r,t-1}\sigma(\cdot)$be a sigmoid function, $\phi(\cdot)$be a non-linear activation function, $W_1,W_2,b_1,b_2$and be trainable parameters. Then, the component failure probability is:

$$z_{r,t}=\left[x_{r,t};\Delta T_{r,t};h_{r,t-1}\right],\quad p_{r,t}=\sigma\left(W_2\phi\left(W_1 z_{r,t}+b_1\right)+b_2\right) \tag{4}$$

Equation (4) signifies that image thermal texture, regional statistics, and historical status are all incorporated into the judgment. The model output $p_{r,t}$is not directly equivalent to an alarm, but rather serves as the input to the risk integral model. This can suppress short-term false alarms caused by momentary occlusion, air conditioning switching, and camera noise.

4.4 Multi-level early warning rules and training objectives

To reflect the persistence of anomalies, the system employs a cumulative risk integral. Let be the risk integral of the region $S_{r,t}r$at time t, λbe the historical decay coefficient, and θbe the warning trigger benchmark. Combining the classification cross-entropy, segmentation Dice loss, and safe temperature penalty term, the warning integral and training objective can be obtained:

$$S_{r,t}=\max\left(0,\lambda S_{r,t-1}+p_{r,t}-\theta\right),\quad L=CE(y,\hat{y})+\eta\left(1-Dice(m,\hat{m})\right)+\rho\max\left(0,T_{r,t}^{max}-T_{safe}\right)^2 \tag{5}$$

As shown in equation (5), Sr,t is used to distinguish between instantaneous anomalies and continuous anomalies; CE is the classification loss; Dice measures the degree of overlap between the actual thermal defect mask and the predicted mask; Tsafe is the equipment safety temperature reference value; and eta and rho are weighting parameters. The four risk levels in the system are normal, attention, warning and alarm. The degree is not solely determined by the model's probability, but also by the joint satisfaction of maximum temperature, load-corrected temperature rise, continuous duration and spatial consistency of the hot spot.

Table 3. Task-specific Comparison of Infrared Thermal Imaging Fault Detection Methods in the Past Three Years

Ref.	Published scenario	Core method	Task-specific metric	Data center adaptation issue
[3]	Server monitoring	CNN thermal diagnosis	Classification: Accuracy >98%	Need more rack types and live fault cases
[4]	Complex electrical equipment	Infrared recognition	Classification: Accuracy 86.54%-90.91%; FAR 7.50%-8.20%	Rack airflow and occlusion not covered
[5]	Substation equipment	Lightweight CNN	Classification: high classification performance	Device shape differs from PDU and UPS parts
[6]	Power equipment	PEDNet YOLOv4-Tiny	Detection: lightweight target detection	Small plugs and cables need special labels
[7]	Substation objects	Improved YOLOv8	Detection: multi-scale detection improvement	Dense cabinet background needs retraining
[8]	Conveyor idlers	U-Net segmentation	Segmentation: mean Jaccard 93.86%	Industrial texture differs from cabinets
[9]	Motor and gearbox	CNN diagnosis	Classification: Accuracy 98.54%; F1 98.55%	Mechanical categories differ from contacts
[10]	Cable terminal	Fusion-YOLOv5	Detection: mAP@0.5 95.3%; 12 ms	Terminal geometry differs from PDU modules

As shown in Table 3, the infrared thermal imaging fault detection methods in the past three years include classification, detection and segmentation. The reported values are organised by task-specific indicators and are not normalised into a single score. For data center deployment, Accuracy and F1 score are suitable classification indicators for equipment status judgment; mAP and latency are suitable detection indicators for component location; and Dice or Jaccard indicators are appropriate segmentation indicators for hotspot boundary extraction. The above values can help select a method, but they should not be used directly for cross-task ranking.

4.5 Alarm Level Criteria and Engineering Trigger Standards

Table 4 shows the engineering trigger standards employed by the system proposed above. Delta Tcorr is the load-corrected temperature rise of a component ROI relative to its healthy baseline, and Tmax is the maximum temperature in the ROI. Thresholds are used in combination with duration filtering; thus, short-term airflow disturbances, reflections and temporary occlusions are classified as pending evidence rather than being immediately reported. High-value load areas can use a lower Warning threshold, and frequently plugged-in areas can extend the duration window to avoid false alarms.

Table 4: Multi-level Alarm Criteria for Temperature Rise and Duration

Alarm grade	Temperature criterion	Duration criterion	System action	Reset condition
Normal	Delta Tcorr <5 C and Tmax <45 C	No continuous exceedance for 10 min	Routine recording and baseline update	Maintain normal state
Attention	5 C <= Delta Tcorr <8 C or 45 C <= Tmax <55 C	Continuous for >=10 min	Edge observation; check load and airflow records	Return below 5 C for 15 min
Warning	8 C <= Delta Tcorr <15 C or 55 C <= Tmax <65 C	Continuous for >=5 min	Generate review task and evidence pack within 30 min	Return below 8 C for 20 min after confirmation
Alarm	Delta Tcorr >=15 C or Tmax >=65 C; Delta Tcorr >=20 C triggers immediate alarm	Continuous for >=3 min, or immediate for severe rise	Notify on-site inspection, load transfer or shutdown decision	Manual confirmation and temperature stable below 8 C for 30 min

5. Simulation Case Verification and Operation Evaluation

5.1 Simulated Rack-Aisle Case and Verification Setting

A simulated verification case is built for a 20-rack high-density aisle. The monitored objects are the rear PDU sockets, UPS input/output terminals and local busbar connections. The sampling interval of the evaluation curve is 1 minute, the ambient temperature is around 22.5℃, and rack load variations are added to verify whether the normal load heating can be separated from abnormal contact heating. Three disturbances are to be created: a normal rack-load increase, a short-term hot spot caused by airflow obstruction, and a sustained PDU terminal loosening fault. Table 4 and the risk integral in Equation (5) are used in the evaluation logic, not a single-frame model output.

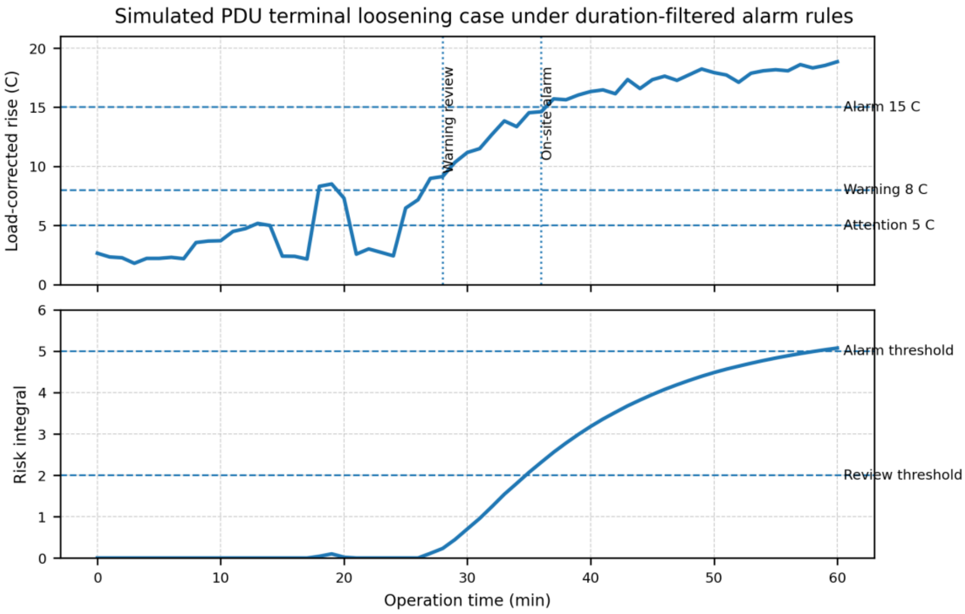

Figure 3. Simulated PDU Terminal Loosening Case under Duration-filtered Alarm Rules.

Figure 3 is the simulated temperature-rise curve and the cumulative risk integral. A short-lived hot spot around 18-20 minutes exceeds the Warning Temperature-Rise Line but does not meet the duration requirement, so it is excluded. After 25 minutes, the PDU terminal loosening fault causes a continuous rise in load-corrected temperature. The system will create a warning task when the conditions of the warning threshold and duration are met, and trigger an on-site alarm after the risk integral reaches the alarm threshold. Therefore, a persistence filter can be applied to reduce temporary false alarms and maintain sensitivity to extended high-temperature conditions.

Table 5. Simulation Case Verification Results

Case event	Maximum Delta Tcorr	Duration over threshold	System result
Normal load increase	5.1 C	Below Attention duration	Recorded as load-related heating; no alarm
Temporary airflow obstruction	8.4 C	3 min	Filtered as pending evidence; no work order
PDU terminal loosening	18.9 C	Over Warning for 32 min; over Alarm for 24 min	Warning review at 28 min; on-site Alarm at 36 min

Based on the case results, the proposed multi-level rule can distinguish between load fluctuation, short-term thermal interference and prolonged fault heating. A fixed threshold strategy retains image evidence of ambiguous events and only issues an executable work order when there is consistency in temperature rise, duration and asset location.

5.2 Implementation Path and Operation Evaluation

The recommended system deployment path is "pilot area-grey-scale operation-closed-loop optimisation". The first stage selects UPS input/output terminals, PDU sockets and the rear area of high-power cabinets as pilot sites to build asset ROI templates, healthy temperature baselines and alarm thresholds. The second stage is an edge-side greyscale operation; only when the duration and space-consistency conditions are met will a formal work order be issued. The third stage connects the alert level to DCIM, BMS or ITSM systems and publishes evidence images, component locations, temperature-rise curves, duration records and handling recommendations.

Based on the system requirements, the infrared camera needs to have a stable frame rate and a low NETD; edge nodes should support TensorRT or ONNX Runtime; and the data platform must store raw thermal images, ROI features, baseline parameters, alarm duration and work-order feedback. The software modules are as follows: Camera Management, Thermal Image Calibration, ROI Registration, Model Inference, Risk Integration, Alarm Configuration, Access-Control Auditing and Sample Feedback. To reduce false alarms, the system has time verification, space verification and operation verification: a hot spot must be prolonged, overlap with a component boundary, and deviate from the load baseline before it is flagged for elevation.

Operating analysis includes task-specific algorithm indicators and operating indicators. Classification performance is shown as Accuracy, Recall and F1 score; target detection is shown by mAP and single-frame latency; hotspot segmentation is shown by Dice or Jaccard; and these indicators cannot be directly compared across task types. Operation indicators are the mean time to alarm confirmation, false alarm rate, review pass rate, early fault detection time, and work-order closure time. Group the risk score thresholds by equipment type and set up different safety bases for UPS terminals, PDU sockets and server rear ventilation openings. High-value load areas can use a lower warning threshold for higher sensitivity, and frequently plugged-and-unplugged areas may have a longer persistence window to prevent false alarms.

6. Conclusion

Infrared thermal imaging fault detection and multi-level early warning system for hidden overheating risks in data centre cabinets, PDUs and UPS equipment have been designed. The integrated loop includes infrared thermal imaging, visible-light assisted positioning, environmental parameters, load information, risk integration and maintenance work orders. The revised scheme specifies task-oriented literature metrics, sets concrete alarm grades based on load-corrected temperature rise and duration windows, and introduces a simulated rack-aisle case to verify that transient hot spots can be filtered while sustained PDU terminal loosening triggers an executable alarm. The system can reduce the problems of manual inspection and single fixed-threshold judgment, but large-scale continuous verification in production data centers is still required. Future work will include cross-data centre transfer learning, small-sample fault detection, multimodal state fusion, low-power edge deployment and optimisation of DCIM linkage strategy.

References

[1] Zhao R., Du Y., Yang X., Zhou Z., Wang W., Yang X. A critical review on the thermal management of data center for local hotspot elimination. Energy and Buildings, 2023, 297: 113486. DOI: 10.1016/j.enbuild.2023.113486.

[2] Chen X., Tu R., Li M., Yang X., et al. Hot spot temperature prediction and operating parameter estimation of racks in data center using machine learning algorithms based on simulation data. Building Simulation, 2023, 16: 2159-2176. DOI: 10.1007/s12273-023-1022-4.

[3] Wiysobunri B. N., Erden H. S., Toreyin B. U. CNN-based server state monitoring and fault diagnosis using infrared thermal images. Soft Computing, 2024. DOI: 10.1007/s00500-024-09792-y.

[4] Tang Z., Jian X. Thermal fault diagnosis of complex electrical equipment based on infrared image recognition. Scientific Reports, 2024, 14: 5547. DOI: 10.1038/s41598-024-56142-x.

[5] Zhou S., Liu J., Fan X., Fu Q., Goh H. H. Thermal fault diagnosis of electrical equipment in substations using lightweight convolutional neural network. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 5005709. DOI: 10.1109/TIM.2023.3240210.

[6] Li J., Xu Y., Nie K., Cao B., Zuo S., Zhu J. PEDNet: A lightweight detection network of power equipment in infrared image based on YOLOv4-Tiny. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 5004312. DOI: 10.1109/TIM.2023.3235416.

[7] Xiang S., Chang Z., Liu X., Luo L., Mao Y., Du X., Li B., Zhao Z. Infrared image object detection algorithm for substation equipment based on improved YOLOv8. Energies, 2024, 17(17): 4359. DOI: 10.3390/en17174359.

[8] Siami M., Barszcz T., Wodecki J., et al. Semantic segmentation of thermal defects in belt conveyor idlers using thermal image augmentation and U-Net-based convolutional neural networks. Scientific Reports, 2024, 14: 5748. DOI: 10.1038/s41598-024-55864-2.

[9] Resendiz-Ochoa E., et al. Application of thermography and convolutional neural network to diagnose mechanical faults in induction motors and gearbox wear. Applied System Innovation, 2024, 7(6): 123. DOI: 10.3390/asi7060123.

[10] Yao S., Liu Z., Zhao Y., et al. Automatic diagnosis of heating in oil-filled terminals of cables. Scientific Reports, 2025, 15: 38733. DOI: 10.1038/s41598-025-22506-0.