Engineering Applications of Artificial Intelligence 164 (2026) 113256

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Research paper

FlashDetR: A deep learning pipeline for early detection and time estimation of flashover in high-voltage insulators using infrared videos

Najmath Ottakath [a], Abdullah Lutfi [b], Ali Hamdi [c], Khaled Shaban [a,*], Ayman El-Hag [b]

[a] Qatar University, Al Tarfa Street, Doha, 2713, Qatar
[b] University of Waterloo, 350 Columbia Street West, Waterloo, N2T, Ontario, Canada
[c] MSA University, 6th October City, 12451, Egypt

ARTICLE INFO

Keywords:
Dry band arcing
Flashover
Early prediction
Transformer model
Deep learning
Time-to-flashover

ABSTRACT

Flashover in high-voltage insulators poses a significant risk to power system reliability, potentially leading to outages and safety hazards. This study introduces an innovative deep learning-based approach for early prediction of flashover events and time-to-flashover estimation by analyzing infrared videos of dry band arcing, a known precursor to flashover. In this work, we propose a pipeline named **Flashover Detector and Time Estimator**, which integrates a transformer-based model to accurately predict flashover occurrences, while a Three Dimensional Convolutional Neural Network-based model estimates the time to flashover. Flashover Detector and Time Estimator progressively samples video frames at multiple scales, enhancing prediction accuracy. Experimental results demonstrate that the models achieve up to 88.73% accuracy in predicting flashover events and a mean absolute error of 3.41 in time-to-flashover estimation. These findings substantially improve the ability to implement preventive measures. Flashover Detector and Time Estimator thus represents a significant advancement in proactively managing power system reliability, with demonstrated effectiveness and real-time application potential.

1. Introduction

The efficient functioning of high-voltage power lines is essential for maintaining a continuous supply of electrical energy. Outdoor insulators are critical components of these systems, supporting transmission lines and isolating live conductors from grounded towers (Salem et al., 2022).

Over time, insulators are exposed to environmental contaminants, such as dust and chemical pollutants, which gradually degrade their performance (Cao et al., 2022; Fan et al., 2024). A conductive layer forms on the insulator surface as soluble components in the contamination dissolve, leading to increased leakage current. As the current density on the insulator surface is not uniform, this results in the formation of dry band regions and hence dry band arcing (Lutfi et al., 2022), which can eventually trigger flashovers (Terrab and Bayadi, 2014). Flashover incidents can cause unplanned outages, making accurate flashover prediction vital for ensuring power system reliability and safety (Salem et al., 2022).

Flashover likelihood is influenced by factors such as surface salt density, humidity, temperature variations, and leakage current intensity (Maraaba et al., 2022; Nguyen and Cho, 2025). Monitoring these factors is essential for effective prediction. Infrared (IR) imaging provides a contactless means of monitoring insulator surface conditions. By detecting infrared radiation, IR cameras generate thermograms with pixel values proportional to surface temperature (Wilson et al., 2023).

Elevated surface temperatures indicate contamination. Phenomena such as dry band arcing caused by insulator contamination appear as hot spots that vary with arcing severity, making dry band arcing a key precursor for flashover prediction (Gao et al., 2022).

Traditional image processing methods have been employed to analyze IR images for temperature distribution, as demonstrated in Gao et al. (2022), but these approaches are often reactive. Machine learning (ML) methods, which use features such as leakage current, Equivalent Salt Deposit Density (ESDD), and Non-Soluble Deposit Density (NSDD), have also been explored (Arshad et al., 2020; Mitrovic et al., 2025). However, feature extraction in these methods can be cumbersome. Deep learning offers a more robust solution, enabling automated feature extraction and early flashover prediction by analyzing dry-band arcing patterns directly from video data.

Deep learning, a subset of artificial intelligence, has excelled in image analysis and pattern recognition (Barros and Perez, 2006; Zhao

* Corresponding author.
E-mail addresses: no1912348@qu.edu.qa (N. Ottakath), abdullaeyad007@gmail.com (A. Lutfi), alihamdif@gmail.com (A. Hamdi), Khaled.Shaban@qu.edu.qa (K. Shaban), ahalhaj@uwaterloo.ca (A. El-Hag).

https://doi.org/10.1016/j.engappai.2025.113256
Received 12 March 2025; Received in revised form 17 September 2025; Accepted 17 November 2025
Available online 30 November 2025

et al., 2021; Belhouchet et al., 2024; Orellana et al., 2023). Convolutional Neural Networks (CNNs) and transformer-based models can automatically learn and extract complex features from video data, making them well-suited for identifying and predicting flashover from early indicators (Haj et al., 2021).

Implementing deep learning techniques allows for continuous and automated monitoring of insulators through video analysis. These models can recognize patterns and anomalies that precede flashover events with greater accuracy and speed compared to traditional methods. Furthermore, deep learning can process large datasets, offering a comprehensive and reliable assessment of insulator conditions.

In this study, we introduce FlashDetR (Flashover Detector and Time Estimator) as a novel deep learning pipeline for early flashover prediction and time-to-flashover estimation using IR videos of dry band arcing. FlashDetR leverages progressive action prediction methods derived from transformer-based models (Stergiou and Damen, 2023) to identify early flashover precursors, while a 3D(Three Dimensional) CNN-based model (MoViNet-A4) is fine-tuned for time prediction using a regression approach. This proactive methodology enhances the reliability and safety of power transmission systems, ensuring a stable power supply. To support the development and evaluation of FlashDetR, we curated a novel dataset consisting of 43 infrared (IR) videos of insulators under controlled laboratory conditions. The dataset covers both flashover and non-flashover cases and was designed to reflect realistic surface contamination effects under consistent voltage ramping conditions. To the best of our knowledge, no publicly available dataset exists that provides temporally annotated thermal video sequences of dry band arcing leading to flashover. This lack of comparable resources necessitated the creation of a custom dataset tailored to the specific objectives of early flashover detection and time prediction.

The main contributions of this research are as follows:

- **Development of FlashDetR**: Introduced FlashDetR, a state-of-the-art IR video pipeline specifically designed for early flashover prediction and time estimation of flashover events in high-voltage insulators.
- **Innovative Progressive Prediction Method**: Developed an advanced progressive sampling and multi-scale prediction methodology within FlashDetR, leveraging transformer-based models to enhance early flashover detection accuracy.
- **High Accuracy in Flashover Event Prediction**: Achieved superior performance with up to **88.73**% accuracy in predicting flashover events, surpassing existing methods that rely on manually extracted features.
- **Precise Time-to-Flashover Estimation**: Successfully estimated the time remaining until a flashover occurs with a mean absolute error (MAE) of **3.42**, providing actionable insights for proactive maintenance and intervention.
- **Model Explainability with Grad-CAM**: Integrated Gradient-weighted Class Activation Mapping (Grad-CAM) into FlashDetR to enhance model interpretability, enabling visualization of critical regions within IR videos that influence flashover predictions.
- **Comprehensive Dataset Creation**:
  We curated a specialized dataset comprising **43 infrared (IR) videos**, each featuring a single insulator recorded under fixed and controlled laboratory conditions, covering both flashover and non-flashover scenarios to effectively train and validate the FlashDetR pipeline.
- **Real-Time Application Potential**:FlashDetR demonstrated real-time feasibility with a processing time of **0.4292** seconds, showcasing its scalability for large-scale power networks.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature. Section 3 outlines the methodology followed by the experimental setup. Section 5 presents experimental results and discussion. Finally, Section 6 concludes the paper with a summary of the findings and their implications.

2. Related work

2.1. Prediction of flashover

State-of-the-art flashover prediction methods for high-voltage insulators use both rule-based and ML approaches. Gao et al. (2022) combined temperature probability density from IR images with leakage current data to predict flashover probability. Taibaoui et al. (2022) employed features like insulator geometry and creepage distance in ANN-based models.

Sajjad et al. (2021) used features such as ESDD, NSDD, temperature, and humidity to predict flashover voltage and surface resistance. Zhang et al. (2024b) applied probabilistic neural networks to assess insulator pollution levels, while Zhang et al. (2022) used space electric field signals and Random Forest classifiers to warn against pollution flashovers.

While effective, these methods rely on manual feature extraction and lack real-time capabilities. In contrast, the proposed FlashDetR pipeline automates feature extraction using deep learning on thermal video data, enabling dynamic, real-time early flashover detection by identifying critical patterns of dry band arcing.

Beyond thermal and electrical measurements, ultraviolet (UV) imaging has also been used in the diagnostics of electrical equipment. UV cameras can detect corona discharges and surface defects (Rodrigues et al., 2025). There are no current studies related to flashover prediction using corona discharge. Furthermore, such systems are often sensitive to ambient lighting, require precise calibration, and are less effective for detecting heat-driven phenomena like dry band arcing.

In contrast, the proposed FlashDetR pipeline leverages infrared (IR) thermal video data, which directly captures the heat patterns caused by surface leakage currents and dry band formation—key precursors to flashover. Thermal imaging offers consistent performance under varied lighting and is more suited to monitoring the thermal evolution associated with pollution flashover mechanisms.

2.2. Early prediction

The early prediction of critical events is a vital challenge across various domains, including natural disaster forecasting and industrial accident prevention. Traditional methods have often struggled with timely and accurate forecasting, but advancements in computer vision and machine learning have paved the way for innovative solutions.

In the domain of early action recognition, earlier methods modeled event probabilities in partially observed videos (Cao et al., 2013; Lan et al., 2014). Ryoo (2011) introduced a bag-of-words approach for this purpose, while subsequent studies utilized sparse coding (Cao et al., 2013) and scoring mechanisms (Kong and Fu, 2015). Lan et al. (2014) developed a hierarchical representation of motion in partially observed videos.

Recent approaches have employed advanced techniques such as knowledge distillation to transfer insights from full video observations to partial segments. This has been achieved using Long Short-Term Memory networks (Hu et al., 2018; Pang et al., 2019; Wang et al., 2019), teacher–student configurations (Zheng et al., 2023), and Recurrent Neural Networks with memory augmentation (Cai et al., 2019; Fernando and Herath, 2021). Other notable techniques include Generative Adversarial Networks for feature learning (Xu et al., 2019; Zhang et al., 2024a) and Graph Convolution methods (Liu et al., 2023; Wu et al., 2021b,a). Stergiou and Damen (2023) introduced the Tempr model, which applies temporal scaling to progressively analyze partially observed videos, achieving high accuracy in early action prediction. Leveraging these insights, FlashDetR incorporates a modified Tempr approach specifically tailored to detect early patterns of dry band arcing. By progressively sampling frames, FlashDetR enables timely intervention before flashover occurs, thus mitigating potential system failures.

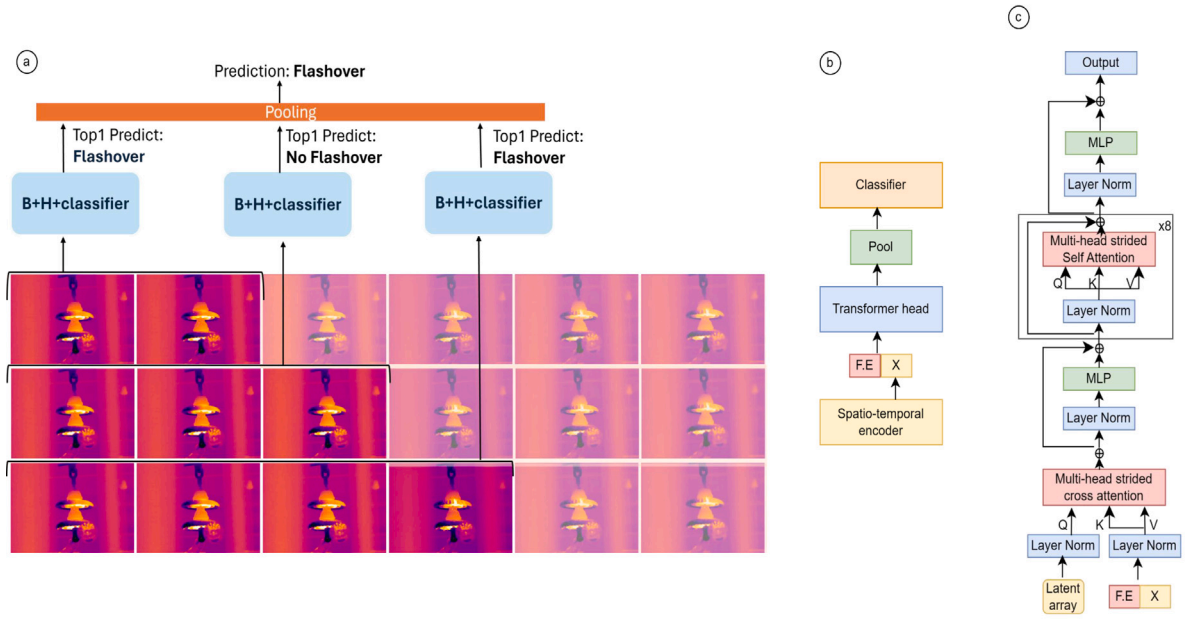

Fig. 1. Tempr model using the infrared dry band arcing video: (a) the overall process, (b) the Tempr model, (c) the transformer head

2.3. Time-to-flashover prediction

While most flashover detection methods focus on classification, there is a critical need to address the prediction of time-to-flashover, which remains largely unexplored. Time-to-flashover prediction offers significant benefits by enabling proactive interventions and optimizing maintenance scheduling. It not only identifies the likelihood of an event but also provides insights into when the event might occur. Time-to-event prediction has been successfully implemented in various domains, such as healthcare for predicting disease progression (Dai et al., 2024) and autonomous systems for collision time estimation (Arrouch et al., 2022; Teng et al., 2024). Inspired by these applications, we extend this concept to flashover prediction.

To achieve time-to-flashover prediction, FlashDetR employs a fine-tuned spatio-temporal model (MoViNet-A4) for video regression. By learning spatio-temporal features from thermal videos, this model provides a continuous estimate the remaining time until a flashover event is likely to occur. This integrated approach within FlashDetR—combining event classification with precise time prediction offers a comprehensive solution for flashover prevention, enhancing both the reliability and the safety of high-voltage power systems.

3. Methodology

3.1. Early prediction of flashover

The methodology for early prediction of flashover in this study involves sampling video frames at multiple scales and feeding these samples into a transformer-based model to generate predictions. The predictions from different scales are then aggregated to produce a final prediction.

This approach builds upon theories from early action prediction, where the goal is to predict a label y for a full video $v(1, 2, \ldots, T)$ based on partial observations. Here, T represents the total number of frames, and the observation ratio (r) (where $0 < r < 1$) determines the number of observed frames during training, $T_r = r.T$. The aim is to predict y using only T_r observed frames. The observation ratio is defined as the fraction of the video that is made visible to the model at training and inference time, while the remainder of the video remains hidden. For instance, an observation ratio of 0.6 means that only the first 60% of frames are available for prediction, simulating an early warning scenario where only partial temporal information is accessible. This setting reflects real-time monitoring conditions, where full future context is not yet available.

Unlike traditional early prediction methods, this approach employs progressive sampling, which generates multiple scales $(K_1, K_2 \ldots, K_n)$ from the observed sequence. Each scale K_i spans a broader temporal extent than the previous scale, capturing the progression of the action from fine to coarse granularity. Frames for each scale are sampled based on the ratio $\frac{i}{n}$ of the total observed frames T_r. Thus, K_i forms a sequence of frames from 1 to T_{Ki}, which is calculated based on the ratio $\frac{i}{n}$ of the total observed frames T_r. Fig. 1 provides an overview of this methodology.

In our implementation, we set $n = 3$, generating three progressively coarser temporal scales. This configuration was chosen after exploring different values for n in the range of 1 to 4. The selected value ensures a balanced representation of the temporal dynamics, allowing the model to capture both fine-grained and coarse-level patterns of the observed sequence.Algorithm 1 represents the temporal scaling approach

Algorithm 1 Progressive Multi-Scale Sampling from Observed Sequence

Require: Video sequence $v = \{v_1, v_2, \ldots, v_T\}$ of length T
Require: Observation ratio $r \in (0, 1]$
Require: Number of scales n, frames per scale L
Ensure: Multi-scale sampled frame sets $\{K_1, K_2, \ldots, K_n\}$
$T_r \leftarrow \lfloor r \cdot T \rfloor$ ▷ Compute number of observed frames
$v_r \leftarrow \{v_1, v_2, \ldots, v_{T_r}\}$ ▷ Observed portion of the video
for $i = 1$ to n **do**
  $T_{K_i} \leftarrow \lfloor \frac{i}{n} \cdot T_r \rfloor$ ▷ Temporal extent for scale K_i
  $S_i \leftarrow \{v_1, v_2, \ldots, v_{T_{K_i}}\}$ ▷ Temporal window up to T_{K_i}
  $K_i \leftarrow \text{RandomSample}(S_i, L)$ ▷ Randomly sample L frames
end for
return $\{K_1, K_2, \ldots, K_n\}$ ▷ Multi-scale inputs for model

At each scale, a fixed number F of frames is randomly sampled to standardize the encoder input. Features are extracted using a shared encoder and embedded with positional information through Fourier positional embeddings. Multiple attention towers process these features at each scale, as shown in Fig. 1(c). Each tower consists of strided cross-attention (Cross-MAB) and stacked strided self-attention (Self-MAB) blocks to refine feature representation.

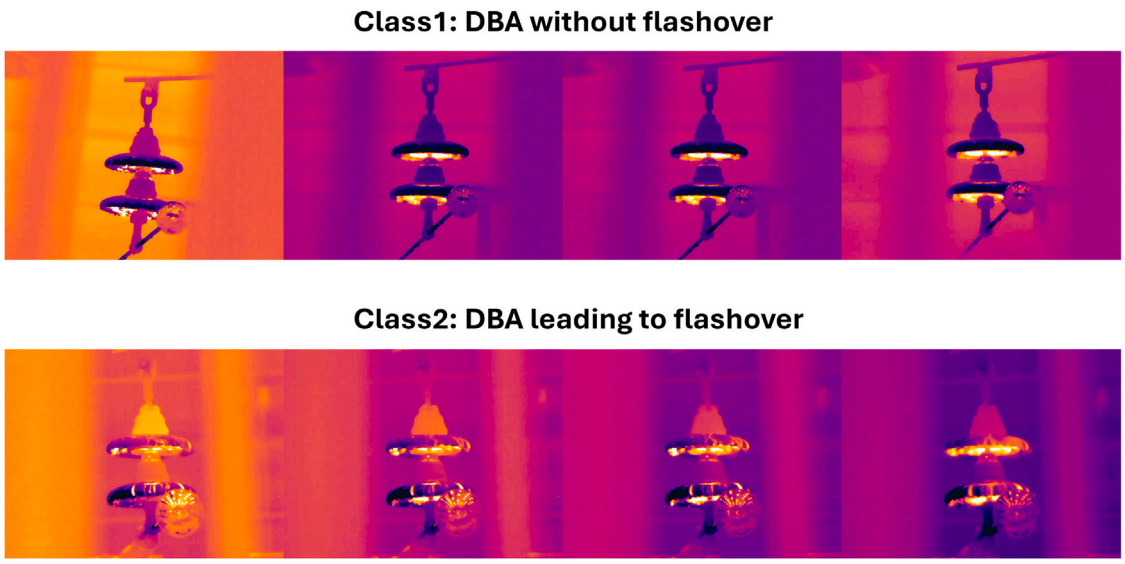

Fig. 2. Progressive frames from the dataset depicting Class 1: Dry band arcing only (no flashover) and Class 2: Dry band arcing leading to flashover (flashover).

Predictions are generated for each scale using a linear classifier, and the final prediction is obtained by aggregating these predictions through an aggregation function that considers two key factors: predictor agreement and predictor confidence.

- **Predictor agreement** is calculated using the Exponential Inverse Coefficient Weighting (eICW) method, which weights predictions based on their similarity to the mean probability distribution y_{bi}. Similarity is measured via the Dice-Sørensen coefficient (DSC).
- **Predictor confidence** is determined using the exponential maximum (eM) function, applying a softmax operation to emphasize high-confidence predictions.

This aggregated prediction, referred to as adaPool (Stergiou and Poppe, 2022), integrates both the similarity among predictions across scales and each predictor's confidence, resulting in a robust and reliable final-class label for the video sequence. The final aggregation function combines the agreement and confidence measures across all scales. The relative importance of each factor is controlled by a learnable parameter β where $0 \leq \beta \leq 1$

Incorporation into FlashDetR: We integrate the above multi-scale prediction strategy into our proposed pipeline, FlashDetR, which orchestrates the transformer-based model and the progressive sampling mechanism. Through FlashDetR, the final aggregated prediction is continuously updated as more frames are observed, enabling real-time early flashover detection.

Quantization aware training: To improve inference efficiency, we applied Quantization-Aware Training (QAT) using PyTorch. During training, the model was augmented with fake quantization modules to simulate the effects of INT8 arithmetic for both weights and activations. This enabled the model to adapt to quantization-induced noise, preserving accuracy after conversion. After training, the model was converted to a fully quantized version. This results in reduced memory footprint and faster inference, with minimal degradation in performance.

3.2. Time-to-flashover

Time-to-flashover refers to the estimated time remaining before a flashover occurs. For a video of length T_video and frame rate F_fps, the total number of frames is:

$$N_{\text{frames}} = T_{\text{video}} \times F_{\text{fps}} \tag{1}$$

Each frame corresponds to a specific timestamp $t_{\text{frame,i}} = \frac{i}{F_{\text{fps}}}, \quad i = 0, 1, 2, \ldots, N_{\text{frames}} - 1$, where i is the frame index. The time-to-flashover for each frame is:

$$t_{\text{flashover,i}} = T_{\text{video}} - t_{\text{frame,i}} \tag{2}$$

For a sequence of sampled frames starting at Sj with n frames, the time-to-flashover is:

$$t_{\text{flashover,Sj}} = T_{\text{video}} - t_{\text{frame,j+n-1}} \tag{3}$$

To predict time-to-flashover, the MoViNet architecture is employed due to its strong spatio-temporal features extraction capabilities. Originally designed for classification (Kondratyuk et al., 2021), MoViNet was modified for regression by replacing its classification layer with a fully connected layer that outputs a single scalar value. The loss function used is mean squared error (MSE), minimizing the prediction error for continuous values. MoViNet includes 3D convolutional layers, squeeze-and-excite blocks, residual connections, and 3D pooling layers, which together dynamic content across both spatial and temporal dimensions.

Fine-tuning of the MoViNet is performed where, a selective layer-freezing strategy is adopted to balance pre-trained knowledge retention with domain-specific adaptation. Freezing early layers—such as the initial convolution and the first few backbone blocks—helps preserve low-level spatio-temporal representations learned from large-scale video datasets, including texture, motion, and illumination patterns. These features remain relevant in dry band arcing analysis, where visual cues such as arc flicker, and brightness changes exhibit structured temporal behavior. This strategy also mitigates catastrophic forgetting, reduces overfitting on the limited dataset size, and enhances computational efficiency by reducing the number of trainable parameters.

To identify the most effective layer-freezing configuration, an empirical analysis was conducted by systematically evaluating multiple combinations of frozen layers. This iterative experimentation involved progressively freezing the initial convolutional layer and subsequent backbone blocks, allowing for a controlled assessment of how each configuration influenced regression performance. This empirical approach is justified by the fact that the optimal layer-freezing strategy can vary depending on the complexity of the task, the nature of the target domain, and the distribution shift between the pre-training and fine-tuning datasets. Since no universally optimal configuration exists across tasks, systematically evaluating different layer combinations provides practical insight into how much of the pre-trained knowledge should be retained versus adapted. This enables a data-driven calibration

of model flexibility, ensuring the architecture is neither under- nor over-constrained during fine-tuning.

Within FlashDetR, the MoViNet-based regression module continuously estimates the remaining time to flashover. By coupling this time-to-event prediction with the early classification output, FlashDetR not only identifies whether a flashover is likely but also determines when it might occur, allowing operators to take preventive action.

3.3. Grad-CAM for regression

To interpret model predictions and enhance explainability, we employ Gradient-weighted Class Activation Mapping (Grad-CAM) to identify the most influential regions within each video frame that contribute to the predicted time-to-flashover values. This visualization technique provides crucial insights into the model's decision-making process by generating spatial attention maps that highlight which areas of the insulator surface the model prioritizes during dry band arcing progression analysis.

The Grad-CAM implementation follows a systematic approach: first, we capture feature map activations and their corresponding gradients during the model's forward pass through the convolutional layers. Next, we compute importance weights by performing global average pooling on the gradients, which quantifies how much each feature map contributes to the regression output. These weights are then applied to create a weighted combination of the activation maps, followed by ReLU activation to retain only positive influences and normalization to produce interpretable heatmaps.

The resulting visualizations reveal spatially localized patterns in dry band formation and arcing behavior that correlate with flashover timing, enabling power system engineers to understand not only when flashover might occur, but also which specific regions of the insulator surface drive these predictions. This interpretability is essential for building trust in the model's outputs and facilitating its adoption in critical power system applications where understanding the reasoning behind predictions is as important as the predictions themselves.

4. Experimental setup

4.1. Dataset

The dataset used in this study consists of 43 IR videos, including 22 videos capturing dry band arcing leading to flashover and 21 videos of dry band arcing without subsequent flashover. Each video has a resolution of 640 × 480 pixels. Fig. 2 presents sample frames from the dataset, illustrating the progressive development of dry band arcing and its potential to either result in flashover or not. Notably, areas exhibiting high-temperature spots are significantly larger in cases that lead to flashover compared to those that do not. The time to flashover dataset contains 22 videos of which the average time to flashover was 32.67 s, with a maximum of 45.13 s and a minimum of 18.33 s, across a total of 22 videos. These values represent the time from the earliest visible frame to the flashover event and reflect the range of conditions captured in the dataset.

Data acquisition: The dataset was acquired in a controlled laboratory environment. A 150 kV/20 kVA test transformer was used to generate and apply the high voltage required for the experiment. The test object consisted of a string of two porcelain disc insulators arranged in series, as depicted in Fig. 3, which provides a detailed view of the experimental setup. This configuration was selected to simulate real-world conditions and facilitate the collection of data essential for the development of an AI-based flashover prediction model. The insulators were contaminated to a medium severity level, in accordance with IEC 60507 standards (Internationale, 2013) . Prior to contamination, the insulators were thoroughly washed with distilled water and left to air-dry naturally for 24 h. The insulators were contaminated with a salt-water solution corresponding to a medium contamination level, with an estimated surface conductivity of approximately 7.5 S/m. The contaminant was uniformly sprayed over the insulator surface prior to voltage application. Then the voltage is applied and increased gradually. As the voltage increases, dry band arcing starts to develop till either it lead to flashover or the contaminant dries out and the dry band arcing is extinguished. After that, the contaminant is reapplied and the experiment is repeated. The contaminant was applied under lab controlled condition with the temperature of 22 °C

To assess the thermal behavior of the insulators, an infrared (IR) video was captured using a FLIR thermal camera positioned at two meters from the insulator string. All videos were recorded using a fixed laboratory setup. This standardization was essential to ensure consistency across experiments and eliminate variability in the camera's viewpoint, lighting, or distance. At this stage, no data augmentation techniques were applied. The camera used is FLIR T650sc thermal imaging camera that offers a high-resolution 640 x 480 LWIR detector, a temperature range of −40 °C to 2000 °C, and an accuracy of ±1 °C. The sampling frequency is 1 Hz (1 frame per second). This flashover phenomena usually takes seconds to minutes, so, 1 FPS is enough to capture the dry band arcing leading to flashover. The applied voltage was incrementally increased to enable real-time monitoring of the thermal response. For the dry band arcing class, the voltage was increased and maintained at the string's rated voltage (20 kV in this case). Persistent dry band arcing activity was observed until the insulators dried, at which point the arcing gradually diminished. Conversely, for the flashover class, the voltage was continuously ramped up, and the progressive development of DBA was monitored, with the number and intensity of IR hotspots progressively increasing until flashover event occurred. While this controlled setting enables clear analysis of thermal patterns, we acknowledge that it may limit generalizability to field conditions where camera placement and environmental factors can vary. As part of future work, we plan to expand the dataset to include variable viewpoints, distances, and augmented samples to improve the model's robustness in operational environments.

From a thermal perspective, the infrared patterns captured by the camera primarily result from heat generated by surface leakage currents and subsequent dry band arcing. While the contamination level influences the magnitude and rate of heat development, it does not fundamentally alter the spatial pattern of heat distribution on the insulator surface. That is, the characteristic thermal signatures associated with dry band formation remain consistent, regardless of slight variations in contaminant conductivity. Because our model is designed to recognize these spatial patterns prior to flashover, rather than to predict the absolute rate or severity of heating, a single, consistent contamination level suffices to train and validate the pattern recognition capability without introducing extraneous variability. We recognize, however, that outdoor insulators are exposed to a wide spectrum of contamination conditions and weather effects, humidity, wind, precipitation, and so forth that can influence both thermal response and overall insulation performance. Incorporating such variability into our dataset can be incorporated in future work to enhance model robustness and real world applicability.

Within FlashDetR, this dataset serves as the primary resource for training and validating both the early flashover classification and time-to-flashover regression components. The curated IR videos enable FlashDetR to learn distinct thermal signatures and progression patterns, which are critical for proactive flashover detection and timing.

4.2. Training parameters

The early prediction model was evaluated across four distinct scales. It utilized a latent bottleneck dimension of 256, with four cross-attention heads and eight self-attention heads. For each scale, 16 frames were sampled. Spatio-temporal encoding was achieved using a pre-trained MoViNet model (Kondratyuk et al., 2021), originally trained on the Kinectics-800 dataset.

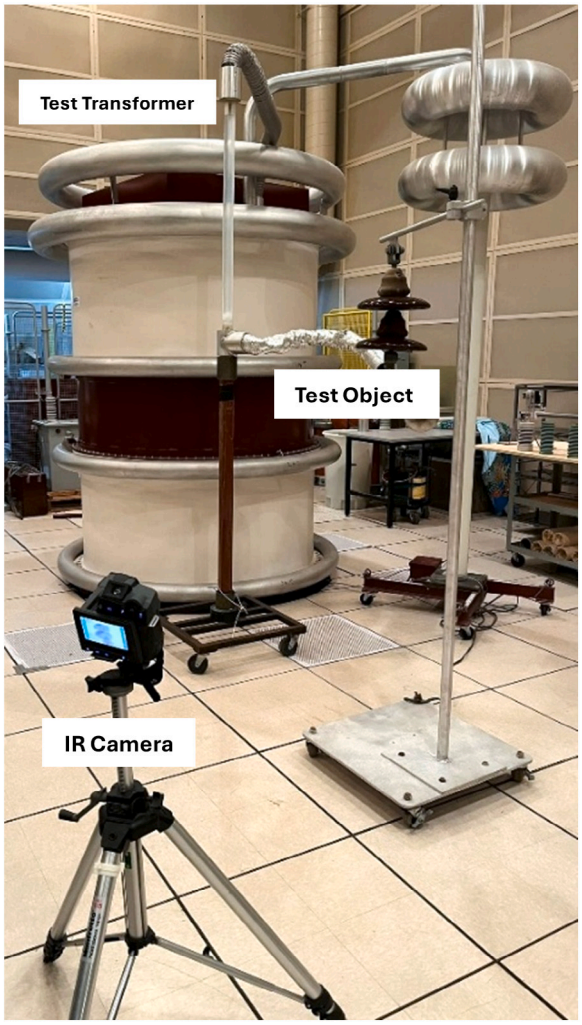

Fig. 3. Data acquisition setup.

The parameter β was initialized to 0.5. Training for early prediction was conducted for 50 epochs on a Google Colab T4 GPU, with a learning rate of $1e{-}2$ for the modified Tempr head and $1e{-}3$ for β. Time-to-flashover prediction was trained for 20 epochs with early stopping (patience of 5), and a learning rate reduction triggered after 3 epochs of plateauing. The initial learning rate was set to $1e{-}4$. A batch size of 2 was used for both tasks. Images were resized to 224×224 pixels for training, validation, and testing. Leave-One-Out Validation (LOOV) was employed for evaluation. The training and validation split was 80-20. The optimizer used was adam for both the experiments.

These training parameters were fine-tuned to balance accuracy and inference speed in FlashDetR. Specifically, using four scales helps capture various temporal resolutions necessary for reliable early detection, while the MoViNet backbone provides efficient spatio-temporal representations for predicting both flashover events and the time remaining until flashover.

To mitigate overfitting, several safeguards were implemented. First, we employed Leave-One-Out Validation (LOOV), ensuring that in each fold, the model was tested on completely unseen videos, thus maximizing generalization. Second, early stopping was applied during training based on validation performance, halting training after five epochs without improvement. Third, we used pre-trained backbone models (MoViNet-A4 pre-trained on Kinetics-800), which enabled better feature generalization and reduced the risk of overfitting to the limited dataset. These combined strategies enhance the model's robustness across diverse video samples.

Table 1
Performance at different observation ratios for each scale.

Scale	Observation ratio			
	0.2	0.4	0.6	0.8
1	74.42	65.12	79.07	81.4
2	72.09	83.72	69.77	76.74
3	76.74	81.4	79.07	81.4
4	76.74	79.07	**88.37**	83.72

Table 2
Performance at different observation ratios for each scale and class.

Scale	Class	Observation ratio			
		0.2	0.4	0.6	0.8
1	No Flashover	0.81	0.71	0.81	0.81
	Flashover	0.68	0.59	0.77	0.82
2	No Flashover	0.67	0.86	0.76	0.76
	Flashover	0.77	0.82	0.64	0.77
3	No Flashover	0.76	0.81	0.76	0.76
	Flashover	0.77	0.82	0.82	0.86
4	No Flashover	0.67	0.90	0.86	0.90
	Flashover	0.86	0.68	**0.91**	0.77

5. Results and discussion

5.1. Early prediction of flashover

5.1.1. Performance evaluation across different scales

Table 1 presents the performance of the early flashover prediction model across varying observation ratios for different scales using the multi-scale Tempr-like approach using LOOV training and testing. The observation ratio represents the proportion of frames used for prediction, simulating scenarios where only partial information about the progression towards flashover is available. Highlighted values indicate the highest accuracy per scale.A sequence of 16 frames for each scale were used for training, validation and testing.

The results demonstrate the model's ability to capture relevant features at different scales and observation levels. For instance:

- At scale 1, the highest accuracy of 81.4% is achieved at an observation ratio of 0.8.
- At scale 2, the highest accuracy of 83.72% occurs at an observation ratio of 0.4, indicating robustness in early-stage prediction.
- At scale 3, accuracy stabilizes between 79% and 81.4%, showing consistent performance across various input lengths.
- At scale 4, the model achieves the best overall accuracy of 88.73% at an observation ratio of 0.6, highlighting its capacity to capture discriminative features essential for early prediction.

The fluctuation in accuracy at lower observation ratios suggests challenges in differentiating flashover from no-flashover scenarios when limited data is available. However, the model's peak performance at intermediate observation levels supports the hypothesis that flashover is a cumulative process with increasingly evident patterns over time.

Table 2 shows class-wise accuracy for flashover and no-flashover at various observation ratios and scales, highlighting the model's ability to differentiate the classes with increasing temporal information. Scale 4 achieves the highest flashover accuracy of **91**% at a 0.6 observation ratio and **90**% accuracy for no-flashover at 0.4 and 0.8 ratios. Scale 3 remains consistent, peaking at 86% for flashover at a 0.8 ratio, while Scale 1 improves steadily, reaching 82% accuracy at 0.8.

These results emphasize Scale 4's effectiveness in capturing discriminative temporal features, particularly at intermediate ratios, and Scale 3's reliability across all ratios. The model's performance reflects the inherent challenges of the dry band arcing phenomenon, where intermittent no-arc phases introduce ambiguity

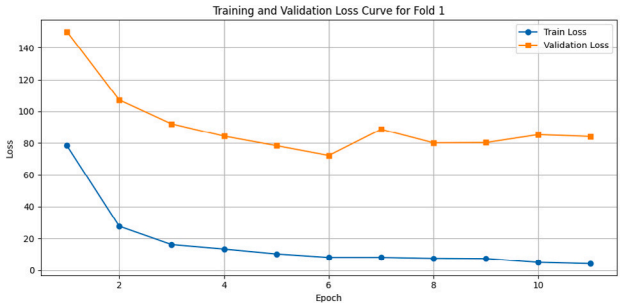

Fig. 4. Training–validation loss curve fold 1.

Table 3
Comparison of Mean MSE and Mean MAE for Different Configurations.

Configuration	Mean MSE	Mean MAE
Transfer learning	50.71	5.32
freeze all	128.73	8.86
freeze conv1,b0	22.46	**3.42**
freeze conv1,b0,b1	20.99	3.73
freeze conv1,b0,b1,b2	46.26	4.90
freeze conv1,b0,b1,b2,b3	35.26	4.49
freeze conv1,b0,b1,b2,b3,b4	56.36	5.83

FlashDetR incorporates advanced temporal feature representation techniques, enhancing the model's ability to capture the nuances of flashover progression. By leveraging multi-scale analysis and progressive sampling, FlashDetR mitigates ambiguities, ultimately boosting predictive performance.

5.2. Time-to-flashover prediction

Time-to-flashover was predicted using the dry band arcing dataset containing sequences that led to flashover. The evaluation utilized the mean MAE and mean MSE metrics, achieved through fine-tuning the MoViNet-A4 model with LOOV. The sequence length for this experiment was empirically chosen as 16.Additionally, an ablation study was conducted by progressively freezing blocks of the model to identify the optimal configuration for time-to-flashover prediction. Unlike conventional RUL estimation frameworks, our model focuses on the short-term prediction of flashover event rather than the overall lifespan of the outdoor insulator. As such, the predictions represent the time-to-flashover rather than the remaining useful life in a maintenance planning context. The results as shown in Table 3 indicate that freezing only the initial convolutional layer and block_b0 yielded the best performance, achieving a low mean MSE of **22.46** and mean MAE of **3.422**. This configuration demonstrated an optimal balance between leveraging pre-trained features and adapting to the specific nuances of the dataset, effectively capturing spatio-temporal characteristics crucial for accurate predictions. In practice, even 1–2 s of advance warning can enable effective intervention via automated protection systems. With an average MAE of 3.42 s, our model offers a conservative and actionable prediction window in the majority of cases. The Fig. 4 illustrates the training and validation loss curve. It shows that the training loss reduces till it plateaus enabling prevention of overfitting problem.

5.2.1. Effect of fine-tuning

Table 3 compares the impact of freezing layers during fine-tuning on time-to-flashover prediction. Multiple configurations were evaluated, progressively freezing layers from the initial convolutional layer to subsequent blocks. The results indicate that transfer learning, no layers frozen, provides an effective starting point. However, the decision of which layers to freeze during fine-tuning significantly affects performance.

Freezing fewer layers allows the model to retain general features learned during pre-training while adapting deeper layers to task-specific data. The optimal configuration involved freezing only the initial convolutional layer and block_b0, which resulted in a mean MSE of **22.46** and a mean MAE of **3.422**, striking a balance between leveraging pre-trained features and adapting to the nuances of the dataset. Conversely, freezing too many layers, such as freezing all except the final fully connected layer, restricted the model's adaptability and led to suboptimal performance, with a high mean MSE of **128.726** and a mean MAE of **8.86**. As more layers were frozen, performance degraded further; for instance, freezing up to block_b2 increased the mean MSE to **46.26** and the mean MAE to **4.90**.

5.2.2. Evaluation using LOOV

To evaluate time-to-flashover prediction, LOOV was performed, training on the dataset while leaving one video out for testing. Fig. 5 presents regression plots for LOOV folds, with each plot corresponding to an excluded video. Blue points show actual vs. predicted times, the green diagonal line represents perfect predictions, and the red line indicates the trend between predicted and actual values. Closer alignment to the green line signifies better performance.

The model's accuracy varied across videos, performing well for MOV_9247, MOV_9239, MOV_9234, and MOV_9230 but showing deviations for MOV_9223, MOV_9240, and MOV_9229. These variations highlight the need for robust feature extraction and adaptive learning within the FlashDetR pipeline to handle diverse flashover scenarios effectively.

In the best-performing model (MOV_9247), predictions closely align with ground truth, indicating minimal deviations and robust temporal understanding.

In contrast, the worst-performing model (MOV_9240) struggles to capture the overall trend, with accurate dips near the final sequences but poor predictions in earlier sequences, highlighting a failure to generalize temporal features.

Despite some deviations in individual folds, the model demonstrates consistent performance across the majority of videos. Specifically, in 16 out of 22 LOOV folds, the absolute prediction error remained below 6 s, which we consider an acceptable intervention threshold for automated mitigation systems. This tolerance reflects the practical utility of the model even in scenarios with moderate temporal inaccuracies. The larger errors observed in a few cases—such as MOV_9240 may be attributed to intermittent or non-uniform surface contamination, which affects the consistency of dry band arcing and thus reduces the temporal regularity needed for accurate prediction. These inconsistencies can limit the model's ability to extract reliable temporal cues.

These findings emphasize FlashDetR's potential for improvement in handling diverse temporal patterns in dry band arcing.

5.2.3. Model explainability using Grad-CAM

Grad-CAM provides insights into the model's decision-making process by visualizing salient regions across temporal sequences of input frames. The analysis compares low-error (best-performing folds) and high-error (worst-performing folds) across multiple leave-one-out runs.

The results indicate clear differences across model performance levels. High-performing models (Fig. 6a) exhibit consistent and localized attention on physically relevant regions of the insulator, particularly the sheds and arc channels. These activations remain stable over time, allowing the model to track the thermal and structural cues associated with dry-band arcing as it progresses toward flashover. In the corresponding low-error folds, attention maps demonstrate strong temporal coherence, with minimal leakage into irrelevant areas. Conversely, in the worst fold of the same model (Fig. 6b), the attention becomes scattered, extending toward metallic pins and background regions. This diffusion of focus disrupts temporal consistency and corresponds to reduced accuracy.

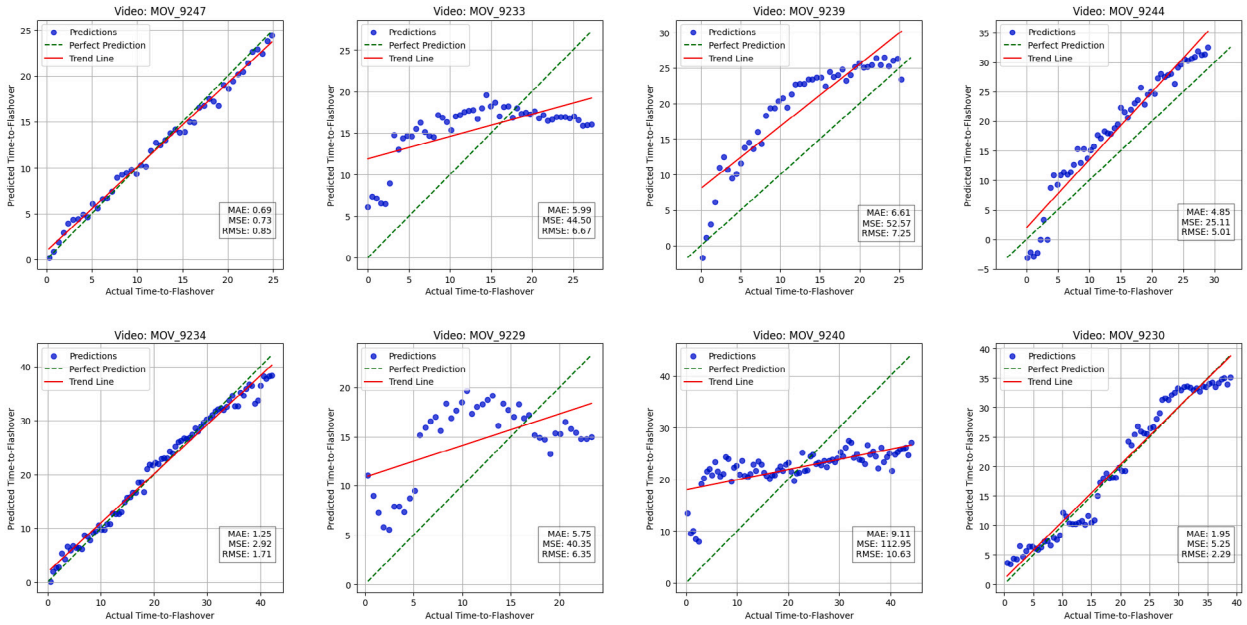

Fig. 5. Regression plots for each LOOV model, showing predicted vs. actual time-to-flashover. Blue points indicate model predictions, the green dashed line denotes perfect predictions, and the red line represents best-fit trends. Key performance metrics—MAE, MSE, and RMSE—are included in each plot, highlighting the alignment of predictions with actual values.

Mid-performing models, shown in Fig. 7, demonstrate partial localization on the insulator body. In the best fold (Fig. 7a), some relevant regions are highlighted, but noticeable portions of the attention still leak into the background. In the worst fold (Fig. 7b), attention frequently shifts away from the insulator, leading to diminished predictive power and unstable regression outputs. This intermediate pattern reflects the limited ability of these models to consistently prioritize physically meaningful regions.

The lowest-performing models, illustrated in Fig. 8, reveal the most significant shortcomings. In the best fold (Fig. 8a), the model captures fragments of the insulator region but continues to attribute relevance to surrounding background areas. In the worst fold (Fig. 8b), critical regions are often missed altogether, with frames showing attention dispersed outside the insulator or even ignoring it entirely. Such failure to maintain coherent focus across frames severely limits the model's ability to capture temporal trends, resulting in poor generalization and inaccurate time-to-flashover estimation.

Overall, the comparison across Figs. 6, 7, and 8 demonstrates that reliable regression performance is strongly correlated with consistent and meaningful attention patterns. Models that maintain focus on the insulator sheds and arc channels produce accurate predictions, while models with dispersed or unstable attention fail to capture essential temporal dynamics. These findings confirm that Grad-CAM not only highlights the regions most relevant to the model but also reveals diagnostic differences between robust and unreliable predictions. As such, Grad-CAM serves as both a validation mechanism and a tool for understanding and improving model behavior, enhancing transparency and fostering expert trust in FlashDetR's predictions.

5.3. Real-time inference capabilities

FlashDetR demonstrated real-time potential with an average inference time of 0.3833 s across four scales, as shown in Table 4. The preprocessing time of a sequence of 16 frames was found to be 0.0110 s. The quantized model further reduced this to 0.3712 s, a gain of 0.0121 s. Additionally, the time-to-flashover model achieved an inference time of 0.0245s per sample, enabling near-instantaneous predictions. With a total processing time of 0.4292 s at scale 4, FlashDetR ensures real-time safety through balanced efficiency and precision.

Table 4
Inference times for the flashover prediction model and its quantized version across different scales.

Scale	Infer time (s)	Infer time quantized model (s)
1	0.3567	0.3522
2	0.3834	0.3572
3	0.3861	0.3706
4	0.4071	0.4047
Avg time:	0.3833	0.3712

5.4. Comparison with related work

FlashDetR demonstrates strong performance in both early flashover prediction and time-to-flashover estimation by leveraging infrared video inputs and deep spatio-temporal modeling. Conventional approaches that rely on manually extracted features—such as leakage current or ESDD—often fail to capture the complex dynamics of dry band arcing and are limited by their need for physical intervention, such as the removal of insulators from the network. These limitations restrict their use in real-time monitoring applications. In contrast, FlashDetR utilizes transformer-based architectures and 3D CNNs for fully automated feature extraction, enabling accurate, real-time predictions directly from thermal video data. Its ability to perform reliably even under partial observations makes it particularly effective for early warning systems in high-voltage network environments. Additionally, we performed extensive validation using 22 leave-one-out folds and conducted ablation experiments to ensure the robustness of our fine-tuned MoViNet backbone.

5.5. Limitations and future work

Despite promising results, FlashDetR has certain limitations:

- **Dataset Size and Diversity:** The dataset of 43 IR videos is relatively small. Outdoor insulators are exposed to a wide spectrum of contamination conditions and weather effects that can influence

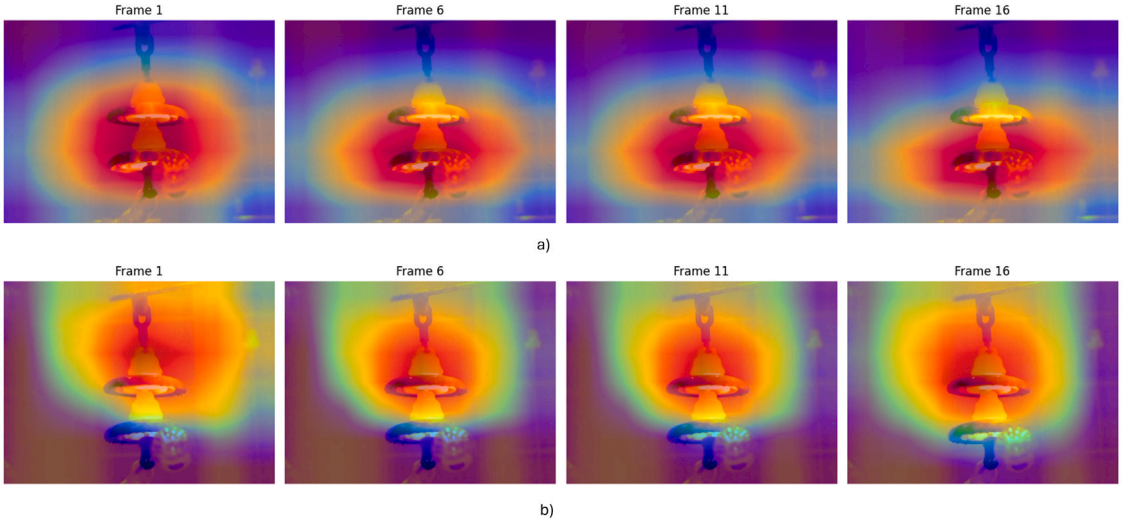

Fig. 6. Grad-CAM for the best-performing model.

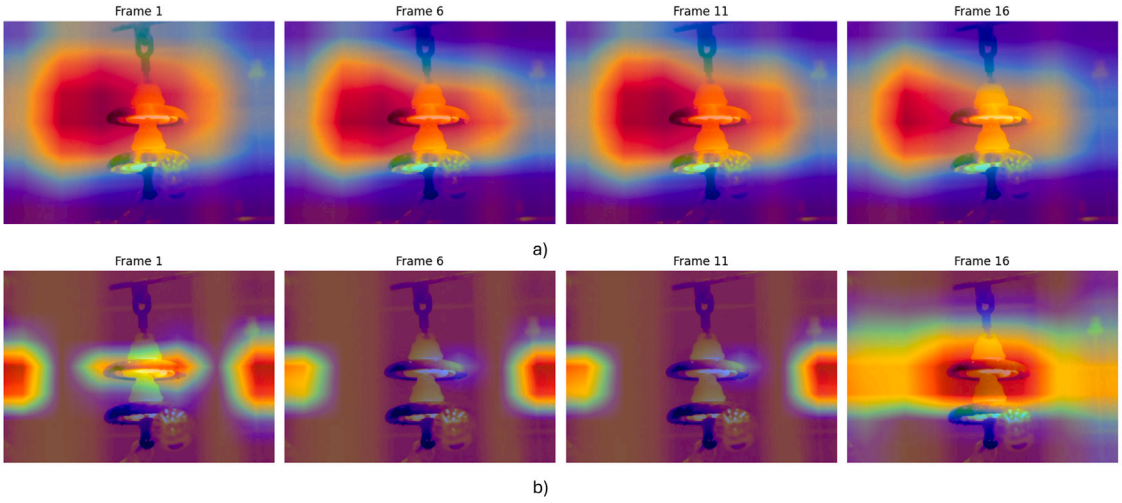

Fig. 7. Grad-CAM for the mid-performing model.

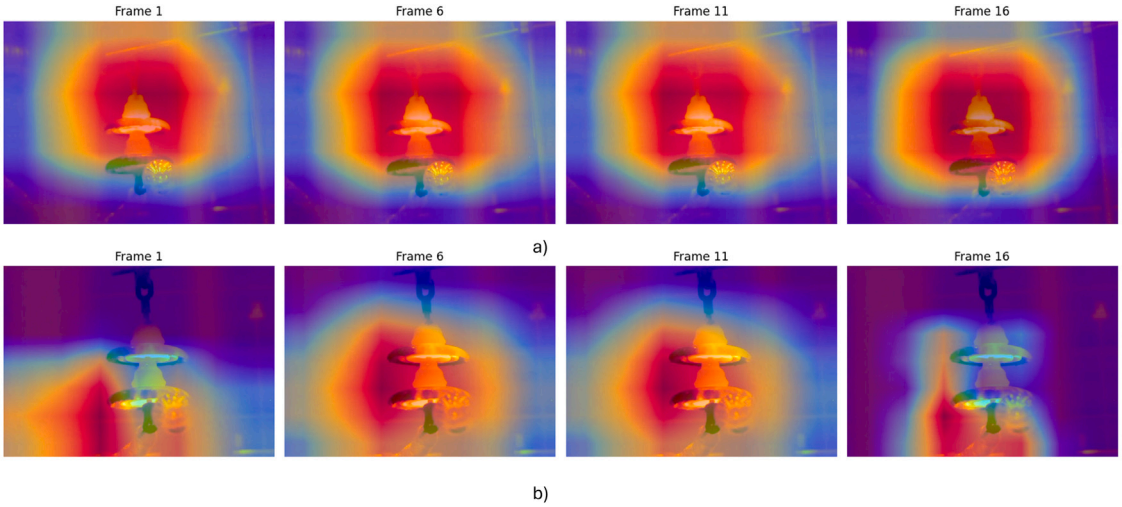

Fig. 8. Grad-CAM for the worst-performing model.

both thermal response and overall insulation performance. Incorporating such variability into our dataset can be done in future work to enhance model robustness and real world applicability.

- **Real-World Deployment:** Lab-acquired data may not reflect real-world complexities—e.g., lighting variations, background noise, sensor inconsistencies. Dry band arcing leading to flashover is typically a long-term process influenced by environmental factors such as humidity, pollution, and insulation degradation over time. To make the phenomenon observable and experimentally manageable within a practical timeframe, we designed an accelerated setup that replicates the underlying physical mechanisms while compressing the timescale. This controlled environment ensured safety, repeatability, and precise labeling of the time-to-flashover, which are challenging to achieve in field conditions. While the dataset does not capture the full range of external variations present in operational settings, it retains the core thermal and spatial patterns of arc evolution. We acknowledge the potential gap in generalizability and plan to address this in future work by incorporating more diverse real-world data or employing domain adaptation techniques to extend model applicability.
  Future research should validate FlashDetR in real-world environments.
- **Model Interpretability:** While Grad-CAM provides valuable spatial visualizations of feature importance, its utility in capturing temporal dynamics is limited, particularly in architectures like MoViNet that process spatio-temporal data. Since the model predicts a continuous target based on sequential visual input, understanding how it weighs temporal progression is crucial. Grad-CAM, being frame-based, may overlook subtle yet critical inter-frame dependencies. To address this, complementary interpretability methods such as saliency mapping over time, temporal attention heatmaps, or model-agnostic tools like LIME and SHAP adapted for video could offer deeper insight into how specific frames or temporal segments influence predictions. Incorporating such techniques can enhance trust in the model's decision-making process and reveal whether the model focuses on physically meaningful aspects of dry band arc development.
- **Real-Time Processing:** Optimizing FlashDetR for edge-device deployment without significant accuracy loss remains a key avenue for future exploration.

Additional data sources (e.g., electrical measurements, environmental sensors) could be integrated to further enhance prediction accuracy and robustness.

6. Conclusion

This study introduced FlashDetR, a deep learning pipeline for the early detection and time estimation of flashover in high-voltage insulators using infrared videos. Leveraging transformer-based models for classification and 3D CNNs for regression, FlashDetR achieved up to 88.73% overall accuracy in flashover prediction and a mean MAE of 3.42 in time-to-flashover estimation. The multi-scale, progressive sampling approach enabled FlashDetR to effectively capture the spatio-temporal dynamics of dry band arcing, providing a robust tool for proactive maintenance and enhanced safety in power systems.

Key contributions of this research include:

- **Development of FlashDetR:** a state-of-the-art IR video pipeline for early flashover detection and time estimation.
- **High Accuracy:** Superior flashover prediction and precise time-to-flashover estimates through advanced deep learning techniques.
- **Model Explainability:** Incorporation of Grad-CAM for interpretability and trustworthiness of predictions.

Future work will focus on expanding the dataset to encompass a wider range of environmental conditions and insulator types, thereby improving the model's generalizability. Furthermore, the incorporation of additional sensor data, such as humidity, temperature, and leakage current, will be explored to enhance prediction accuracy. Another key area to focus will be the scalability of the model for deployment in real-world power systems, ensuring it can effectively handle diverse operational scenarios and maintain performance across varying conditions.

CRediT authorship contribution statement

Najmath Ottakath: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Abdullah Lutfi:** Writing – original draft, Software, Resources, Data curation, Conceptualization. **Ali Hamdi:** Writing – review & editing, Software, Methodology, Conceptualization. **Khaled Shaban:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Ayman El-Hag:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

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

Acknowledgments

This publication was jointly supported by Qatar University grant no. IRCC 554 and NSERC discovery grant no. RGPIN-2020-07004. The findings achieved herein are solely the responsibility of the authors. Open access was funded by Qatar National Library open access fund.

Data availability

Data will be made available on request.

References

Arrouch, I., Ahmad, N.S., Goh, P., Mohamad-Saleh, J., 2022. Close proximity time-to-collision prediction for autonomous robot navigation: an exponential GPR approach. Alex. Eng. J. 61 (12), 11171–11183.

Arshad, Ahmad, J., Tahir, A., Stewart, B.G., Nekahi, A., 2020. Forecasting flashover parameters of polymeric insulators under contaminated conditions using the machine learning technique. Energies 13 (15), 3889.

Barros, J., Perez, E., 2006. Automatic detection and analysis of voltage events in power systems. IEEE Trans. Instrum. Meas. 55 (5), 1487–1493. http://dx.doi.org/10.1109/TIM.2006.881584.

Belhouchet, K., Zemmit, A., Bayadi, A., Ouchen, L., Zorig, A., 2024. A novel application of artificial intelligence technology for outdoor high-voltage composite insulator. Measurement 238, 115372.

Cai, Y., Li, H., Hu, J.-F., Zheng, W.-S., 2019. Action knowledge transfer for action prediction with partial videos. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, (01), pp. 8118–8125.

Cao, Y., Barrett, D., Barbu, A., Narayanaswamy, S., Yu, H., Michaux, A., Lin, Y., Dickinson, S., Mark Siskind, J., Wang, S., 2013. Recognize human activities from partially observed videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2658–2665.

Cao, B., Liu, Y., Li, Z., Shen, S., Wang, L., Wang, Z., 2022. Assessment of insulator pollution degree based on contamination moisture with temperature change. IEEE Sensors J. 22 (21), 21172–21178.

Dai, L., Sheng, B., Chen, T., Wu, Q., Liu, R., Cai, C., Wu, L., Yang, D., Hamzah, H., Liu, Y., et al., 2024. A deep learning system for predicting time to progression of diabetic retinopathy. Nature Med. 30 (2), 584–594.

Fan, Y., Guo, Y., Liu, Y., Xiao, S., Zhang, X., Wu, G., 2024. Evaluation method of composite insulator aging status based on hyperspectral imaging technology. Measurement 225, 113925.

Fernando, B., Herath, S., 2021. Anticipating human actions by correlating past with the future with jaccard similarity measures. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13224–13233.

Gao, K., Ren, M., Jin, H., Tian, H., He, J., Jin, L., 2022. Pollution flashover prediction of insulators based on combined probability density of infrared image temperature and leakage current. In: 2022 IEEE 5th International Electrical and Energy Conference. CIEEC, IEEE, pp. 2014–2019.

Haj, Y.E., El-Hag, A.H., Ghunem, R.A., 2021. Application of deep-learning via transfer learning to evaluate silicone rubber material surface erosion. IEEE Trans. Dielectr. Electr. Insul. 28 (4), 1465–1467. http://dx.doi.org/10.1109/TDEI.2021.009617.

Hu, J.-F., Zheng, W.-S., Ma, L., Wang, G., Lai, J., Zhang, J., 2018. Early action prediction by soft regression. IEEE Trans. Pattern Anal. Mach. Intell. 41 (11), 2568–2583.

Internationale, C.É., 2013. Artificial Pollution Tests on High-voltage Ceramic and Glass Insulators to be Used on Ac Systems. International Electrotechnical Commission.

Kondratyuk, D., Yuan, L., Li, Y., Zhang, L., Tan, M., Brown, M., Gong, B., 2021. Movinets: Mobile video networks for efficient video recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16020–16030.

Kong, Y., Fu, Y., 2015. Max-margin action prediction machine. IEEE Trans. Pattern Anal. Mach. Intell. 38 (9), 1844–1858.

Lan, T., Chen, T.-C., Savarese, S., 2014. A hierarchical representation for future action prediction. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III 13. Springer, pp. 689–704.

Liu, X., Yin, J., Guo, D., Liu, H., 2023. Rich action-semantic consistent knowledge for early action prediction. IEEE Trans. Image Process..

Lutfi, A.E., El-Hag, A., Shaban, K., 2022. Classification of common discharges in outdoor insulators using ultrasonic signals. In: 2022 IEEE Conference on Electrical Insulation and Dielectric Phenomena. CEIDP, pp. 523–526. http://dx.doi.org/10.1109/CEIDP55452.2022.9985259.

Maraaba, L., Al-Soufi, K., Ssennoga, T., Memon, A.M., Worku, M.Y., Alhems, L.M., 2022. Contamination level monitoring techniques for high-voltage insulators: a review. Energies 15 (20), 7656.

Mitrovic, M., Titov, D., Volkhov, K., Lukicheva, I., Kudryavzev, A., Vorobev, P., Li, Q., Terzija, V., 2025. Supervised learning based method for condition monitoring of overhead line insulators using leakage current measurement. Eng. Appl. Artif. Intell. 143, 110040.

Nguyen, T.-P., Cho, M.-Y., 2025. A cloud-based leakage current classified system for high voltage insulators with improved particle swarm optimization and hybrid deep learning technique. Eng. Appl. Artif. Intell. 143, 109987.

Orellana, L., Ardila-Rey, J., Avaria, G., Davis, S., 2023. Danger assessment of the partial discharges temporal evolution on a polluted insulator using UHF measurement and deep learning. Eng. Appl. Artif. Intell. 124, 106573.

Pang, G., Wang, X., Hu, J., Zhang, Q., Zheng, W.-S., 2019. Dbdnet: Learning bi-directional dynamics for early action prediction.. In: IJCAI. pp. 897–903.

Rodrigues, G.A., Araujo, B.V.S., de Oliveira, J.H.P., Xavier, G.V.R., de Souza Lebre, U.D.E., Cordeiro, C.A., Freire, E.O., Ferreira, T.V., 2025. Automated monitoring of insulation by ultraviolet imaging employing deep learning. Measurement 242, 116018.

Ryoo, M.S., 2011. Human activity prediction: Early recognition of ongoing activities from streaming videos. In: 2011 International Conference on Computer Vision. IEEE, pp. 1036–1043.

Sajjad, U., Ahmad, J., Shoaib, S., et al., 2021. Application of artificial neural network in predicting flashover behaviour of outdoor insulators under polluted conditions. In: 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). IEEE, pp. 2868–2873.

Salem, A.A., Lau, K.Y., Rahiman, W., Abdul-Malek, Z., Al-Gailani, S.A., Mohammed, N., Abd Rahman, R., Al-Ameri, S.M., 2022. Pollution flashover voltage of transmission line insulators: Systematic review of experimental works. IEEE Access 10, 10416–10444.

Stergiou, A., Damen, D., 2023. The wisdom of crowds: Temporal progressive attention for early action prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14709–14719.

Stergiou, A., Poppe, R., 2022. Adapool: Exponential adaptive pooling for information-retaining downsampling. IEEE Trans. Image Process. 32, 251–266.

Taibaoui, L., Zegnini, B., Mahdjoubi, A., 2022. An approach to predict flashover voltage on polluted outdoor insulators using ANN. In: 2022 19th International Multi-Conference on Systems, Signals & Devices. SSD, IEEE, pp. 1842–1847.

Teng, X., Xu, S., Guo, D., Guo, Y., Meng, W., Zhang, X., 2024. Dttcnet: Time-to-collision estimation with autonomous emergency braking using multi-scale transformer network. IEEE Trans. Mob. Comput..

Terrab, H., Bayadi, A., 2014. Experimental study using design of experiment of pollution layer effect on insulator performance taking into account the presence of dry bands. IEEE Trans. Dielectr. Electr. Insul. 21 (6), 2486–2495.

Wang, X., Hu, J.-F., Lai, J.-H., Zhang, J., Zheng, W.-S., 2019. Progressive teacher-student learning for early action prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3556–3565.

Wilson, A., Gupta, K.A., Koduru, B.H., Kumar, A., Jha, A., Cenkeramaddi, L.R., 2023. Recent advances in thermal imaging and its applications using machine learning: A review. IEEE Sensors J. 23 (4), 3395–3407.

Wu, X., Wang, R., Hou, J., Lin, H., Luo, J., 2021a. Spatial–temporal relation reasoning for action prediction in videos. Int. J. Comput. Vis. 129 (5), 1484–1505.

Wu, X., Zhao, J., Wang, R., 2021b. Anticipating future relations via graph growing for action prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, (4), pp. 2952–2960.

Xu, W., Yu, J., Miao, Z., Wan, L., Ji, Q., 2019. Prediction-cgan: Human action prediction with conditional generative adversarial networks. In: Proceedings of the 27th ACM International Conference on Multimedia. pp. 611–619.

Zhang, H.-B., Pan, W.-X., Du, J.-X., Lei, Q., Chen, Y., Liu, J.-H., 2024a. Adversarial attention networks for early action recognition. IEEE Trans. Emerg. Top. Comput. Intell..

Zhang, D., Xu, H., Huang, X., Zhang, Z., Jiang, X., 2022. Space electric field characteristics of silicone rubber insulator pollution flashover and its application in flashover prewarning. IEEE Trans. Dielectr. Electr. Insul. 30 (1), 439–448.

Zhang, Z., Zhang, H., Zhou, C., Ma, X., Li, Y., Liu, R., 2024b. A probabilistic neural network assessment method for insulator pollution levels based on infrared images. IEEE Trans. Dielectr. Electr. Insul..

Zhao, Z., Zhang, Q., Yu, X., Sun, C., Wang, S., Yan, R., Chen, X., 2021. Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study. IEEE Trans. Instrum. Meas. 70, 1–28. http://dx.doi.org/10.1109/TIM.2021.3116309.

Zheng, N., Song, X., Su, T., Liu, W., Yan, Y., Nie, L., 2023. Egocentric early action prediction via adversarial knowledge distillation. ACM Trans. Multimed. Comput. Commun. Appl. 19 (2), 1–21.