EviFlash: Uncertainty-Aware Flashover Prediction Using Selective State Space Model

He Xue[1,2], Hang Yin[1], Xiaohan Yang[1], Tingxia Gan[1], Longfei Tan[3], Zhonghai He[4]
{he.xue@scfri.cn, bank2002yinhang@126.com, lamb404@126.com, gantingxia@scfri.cn,
longfei_tan@163.com, hzh@uestc.edu.cn}
[1]Sichuan Fire Science and Technology Research Institute of MEM, Chengdu, 610036, China
[2]Sichuan University, Chengdu, 610065, China
[3]Xihua University, Chengdu, 610039, China
[4]University of Electronic Science and Technology of China, Chengdu, 610054, China

Abstract. Flashover—the abrupt transition of a compartment fire to full development—demands early prediction with quantified confidence for risk-aware operations. We propose an uncertainty-aware evidential flashover predictor (*EviFlash*) that models multi-sensor temperature streams with a selective state space sequence model, enabling long-horizon inference with linear-time complexity suitable for streaming data. The network jointly predicts time-to-flashover (TTF) and a binary flashover-within-horizon event, and represents uncertainty by combining heteroscedastic regression (aleatoric) with stochastic/deep-ensemble inference (epistemic), from which calibrated prediction intervals can be derived. Experiments on two synthetic multi-compartment datasets released by NIST (three- and six-compartment layouts) show that our method consistently reduces TTF mean absolute error and improves recall/F1 over strong recurrent and attention-based baselines, including Bi-LSTM/GRU variants and a recent graph-based flashover model. Ablation studies confirm that both the selective state space backbone and the uncertainty modeling contribute to these gains, and transfer tests across layouts indicate that the proposed model maintains competitive accuracy under changes in compartment geometry and sensor placement, suggesting that uncertainty-aware sequence modeling is a promising direction for early flashover warning in smart firefighting systems without relying on costly computational fluid dynamics (CFD) simulations.

Keywords: Flashover prediction, Selective state space models, Uncertainty quantification, Evidential deep learning, Compartment fires

1 Introduction

Flashover is the abrupt transition of a compartment fire from growth to full development, when radiative feedback from the hot smoke layer rapidly drives most combustibles to ignition [1]. This

IIKI 2025, December 18-21, Chengdu, People's Republic of China

DOI 10.4108/eai.18-12-2025.2365277

transition can occur within seconds and is accompanied by a steep temperature surge, making it one of the most dangerous phases of indoor fires. In dense urban settings, delayed suppression can lead to severe casualties and economic losses; for example, the 1981 Stardust nightclub fire in Ireland escalated rapidly due to flashover, causing 48 deaths and 214 injuries [2]. Flashover is also a major threat to first responders, and has been identified as a leading cause in recent firefighter fatalities and injuries [3]. Reliable early prediction and warning are therefore critical for fire safety operations.

Research on flashover prediction has evolved from analytical criteria and simplified models to high-fidelity simulation and, more recently, data-driven learning. Classical analytical work (e.g., Thomas et al. in the 1970s) used thermal-explosion theory to derive critical conditions for flashover onset [4], but these criteria are typically limited to single-compartment scenarios and struggle with realistic building complexity. Inverse methods estimate heat-release rates and other parameters directly from measurements [5, 6], yet depend strongly on prior knowledge of ignition and ventilation. CFD-based tools such as FDS and CFAST [7, 8] provide detailed spatio-temporal predictions in complex geometries, but their computational cost and sensitivity to model and boundary/material settings restrict large-scale, real-time deployment.

To overcome these limitations, recent work has explored data-driven flashover forecasting using sequence models and spatio-temporal graph networks trained on synthetic sensor data and fire images [9, 10, 11]. These approaches offer fast inference and can generalize across layouts, but they face two key challenges: (i) distribution shift and data scarcity across buildings, fuels, and ventilation regimes, and (ii) the lack of well-calibrated predictive uncertainty, which undermines trust under unseen conditions and complicates risk-aware alarm thresholds.

We address these gaps with **EviFlash**, an uncertainty-aware *evi*dential *flash*over predictor built on a selective state space backbone that captures long-range temporal structure with linear-time inference for streaming multi-sensor signals. EviFlash jointly predicts time-to-flashover (TTF) and flashover-within-horizon events, while decomposing predictive uncertainty into aleatoric and epistemic components via heteroscedastic evidential regression. The resulting prediction intervals support actionable early-warning policies that trade off false alarms and missed detections under operational noise and distribution shift.

This paper makes the following contributions:

- We formulate flashover forecasting as probabilistic sequence modeling on multi-sensor temperature streams, using a selective state space backbone for efficient long-horizon inference.
- We propose an uncertainty quantification scheme that models aleatoric uncertainty via heteroscedastic regression and epistemic uncertainty via stochastic/deep-ensemble inference, yielding calibrated prediction intervals (PIs).
- We demonstrate improved predictive accuracy and classification performance over strong neural baselines on multi-source (synthetic and empirical) datasets, enabling more robust early warning without relying on physics constraints.

2 Related Work

2.1 Flashover modeling and simulation

Early work distilled experimental evidence into simple threshold criteria for imminent flashover, such as upper-layer temperatures of 500°C to 600°C or radiative heat flux above $20\,\mathrm{kW/m^2}$, which are transparent but unreliable in multi-room layouts and complex ventilation. To recover more scene detail from sparse observations, inverse approaches combine simplified physical models with limited measurements to infer latent fire parameters, for example a two-zone model with a genetic algorithm for heat release rate (HRR) estimation [5], a Bayesian inversion framework that yields probabilistic estimates under parameter uncertainty [6], and related numerical treatments across combustion settings [12]. With advances in computation, CFD-based tools such as Fire Dynamics Simulator (FDS) and the Consolidated Model of Fire Growth and Smoke Transport (CFAST) became mainstream for predicting thermal fields and flashover timing in complex geometries [7, 8]. However, these solvers remain computationally expensive and sensitive to model and boundary/material settings, so even with refinements such as probability density functions and state-space reductions [7], there is still a trade-off between physical fidelity and real-time, large-scale deployment.

2.2 Learning-based prediction and uncertainty quantification

Deep learning has enabled data-driven approaches for fire and flashover tasks. Early efforts used convolutional–LSTM surrogates for subgrid flame modeling and random-forest assignment of combustion submodels [13]. Closer to our setting, Lu et al. proposed a CNN–LSTM model that ingests temperature and radiative heat-flux signals from FDS-generated compartment fires with wearable-like sensor placement, achieving reasonable flashover-time errors for lead times up to about 25 s under varying room size, fire load and ventilation [14]. More recent work further exploits temporal models and vision: Song et al. trained ensemble LSTM/GRU models on synthetic temperature and gas measurements plus Vision-Transformer features from fire images, reaching 94–96% accuracy on synthetic data and over 88% on real tests [9]; Mozaffari et al. showed that a transformer-enhanced SwinLSTM on infrared fire video outperforms ConvLSTM on full-scale room fires, supporting vision-based flashover forecasting [11]. Fan et al. introduced gFlashNet, a spatio-temporal graph neural network that represents heat detectors as nodes and edges and uses transfer learning to adapt across residential layouts, achieving high accuracy on unseen layouts while reducing training data and improving scalability [10]. Beyond prediction, Joshi et al. proposed an uncertainty-aware post-detection framework that rescales fire/smoke detections by combining statistical uncertainty with visual cues, improving precision and recall for compact YOLO detectors [15]. These advances underline the growing use of high-capacity models and vision information, but most methods still provide deterministic outputs without calibrated uncertainty for risk-aware decisions.

Modern deep predictors are also known to be miscalibrated, especially under dataset shift. We adopt the standard taxonomy of aleatoric versus epistemic uncertainty and build on techniques such as deep ensembles with data-centric refinements for epistemic modeling [16], conformal prediction for distribution-free coverage [17, 18], and recent studies on calibration principles and conflicting

epistemic losses [19]. In contrast to prior flashover work that largely ignores uncertainty, we target probabilistic sequence modeling with explicit aleatoric and epistemic components, designing an uncertainty-aware state space model tailored to long observation windows and early-warning requirements.

3 Methodology

In this section, we first formalize the flashover forecasting task as probabilistic sequence modeling on multi-sensor time series. We then describe the experimental compartment layouts and sensor configuration, followed by the proposed Mamba-based state space forecaster. Finally, we present our uncertainty quantification design and the overall training objective and early-warning policy. The overall architecture of the proposed framework is illustrated in Fig. 1.

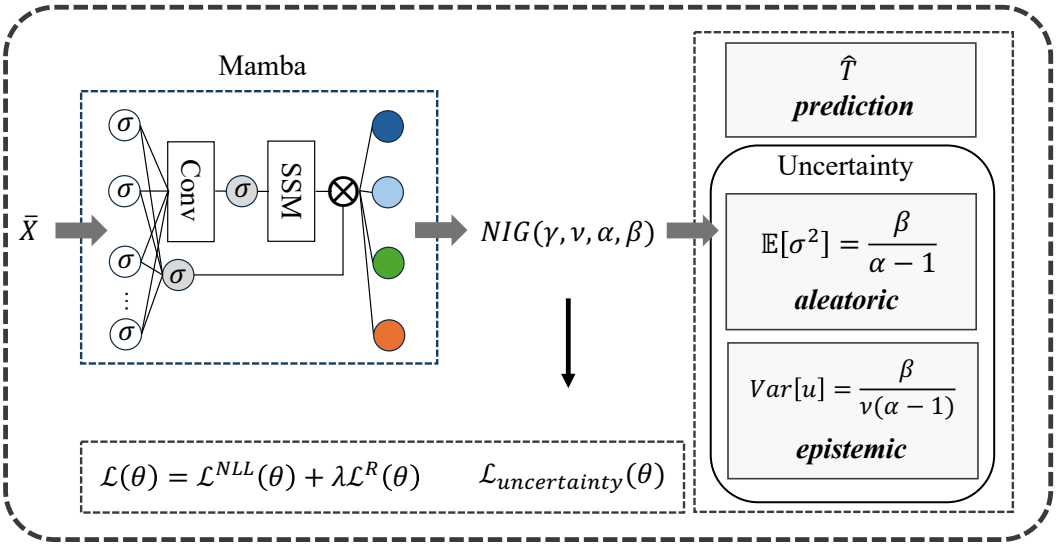

Fig. 1. Overall architecture of the proposed uncertainty-aware flashover prediction framework. Multi-sensor temperature sequences are encoded by a selective SSM backbone, and an evidential head outputs the parameters of a Normal–Inverse-Gamma distribution for uncertainty-aware regression and loss computation.

3.1 Problem formulation

We model flashover forecasting as probabilistic sequence prediction from multi-sensor time series. At each time step t, we observe a d-dimensional vector

$$\mathbf{x}_t \in \mathbb{R}^d \tag{1}$$

obtained by concatenating the temperatures from all heat detectors in the compartment. Given a history window of length T (in experiments, $T = 5$), the input is

$$\mathbf{x}_{1:T} = (\mathbf{x}_1, \ldots, \mathbf{x}_T). \tag{2}$$

For each simulated fire, let t_{fo} denote the time index at which flashover occurs according to the adopted criterion. We define the TTF target associated with this window as

$$y = t_{\mathrm{fo}} - T \tag{3}$$

and a binary event label

$$z = \mathbb{I}\{t_{\mathrm{fo}} - T \leq H\} \tag{4}$$

where H is a predefined decision horizon indicating whether flashover will occur "soon enough" to trigger an alarm.

Our goal is to learn a probabilistic predictor

$$p_\theta(y, z \mid \mathbf{x}_{1:T}) \tag{5}$$

parameterized by θ. From this joint distribution we derive a point estimate and uncertainty for y (TTF), and an event probability for z to support early-warning decisions. The state-space backbone and evidential uncertainty module used to implement p_θ are described next.

3.2 Selective state space backbone

We adopt a selective state space model (SSM) in the spirit of Mamba [20] to encode multi-sensor temperature histories. At each time step t, the raw observation $\mathbf{x}_t$ is first standardized using the training-set statistics,

$$\tilde{\mathbf{x}}_t = (\mathbf{x}_t - u) \oslash \sigma \tag{6}$$

where μ and σ denote the mean and standard deviation and $\oslash$ is element-wise division. A learnable linear layer projects the normalized input to a d_h-dimensional embedding

$$\mathbf{e}_t = \mathbf{W}_{\mathrm{in}}\tilde{\mathbf{x}}_t + \mathbf{b}_{\mathrm{in}} \in \mathbb{R}^{d_h} \tag{7}$$

The selective SSM maintains a hidden state $\mathbf{x}_t \in \mathbb{R}^{d_h}$ and produces an output feature $\mathbf{h}_t \in \mathbb{R}^{d_h}$ at each time step:

$$\mathbf{x}_t = \mathbf{A}_t\mathbf{x}_{t-1} + \mathbf{B}_t\mathbf{e}_t, \qquad \mathbf{h}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{D}\mathbf{e}_t \tag{8}$$

with initial state $\mathbf{x}_0 = \mathbf{0}$. Unlike classical state space models with time-invariant parameters, the transition and projection matrices are generated adaptively from the current input embedding,

$$\mathbf{A}_t = \mathscr{A}(\mathbf{e}_t;\theta), \quad \mathbf{B}_t = \mathscr{B}(\mathbf{e}_t;\theta), \quad \mathbf{C}_t = \mathscr{C}(\mathbf{e}_t;\theta) \tag{9}$$

where $\mathscr{A}(\cdot)$, $\mathscr{B}(\cdot)$ and $\mathscr{C}(\cdot)$ are small neural networks with parameters θ. This *selective* mechanism allows the model to adjust its dynamics at each time step to capture asynchronous sensor behaviour and long-range interactions across compartments.

To obtain a fixed-length representation for the whole observation window, we aggregate the hidden sequence via

$$\mathbf{h}^* = \mathrm{Pool}(\mathbf{h}_{1:T}) \tag{10}$$

where in our implementation $\mathrm{Pool}(\cdot)$ simply takes the last hidden state $\mathbf{h}_T$.

A conventional Mamba regressor would apply a scalar linear head to $\mathbf{h}^*$ to obtain a point estimate of the TTF,

$$\hat{y} = \mathbf{w}^\top \mathbf{h}^* + b \tag{11}$$

To make the backbone uncertainty-aware, we instead replace this one-dimensional head with a multi-dimensional output whose entries are *uncertainty parameters*. Specifically, we compute

$$(\gamma, v, \alpha, \beta) = \mathbf{W}_o \mathbf{h}^* + \mathbf{b}_o \tag{12}$$

and interpret $(\gamma, v, \alpha, \beta)$ as the hyperparameters of a Normal–Inverse-Gamma (NIG) predictive distribution for the TTF target. These parameters are then passed to the uncertainty module (Section 3.3) to obtain both a predictive mean and a decomposition of aleatoric and epistemic uncertainty.

3.3 Evidential uncertainty estimation

Given the sequence representation $\mathbf{h}^*$ from the SSM backbone (Section 3.2), the evidential head outputs the parameters of a Normal–Inverse-Gamma (NIG) distribution for the TTF target:

$$(\gamma, v, \alpha, \beta) = \mathrm{MLP}_{\mathrm{evd}}(\mathbf{h}^*) \qquad v > 0,\ \alpha > 1,\ \beta > 0 \tag{13}$$

Generative model. Following evidential regression [21], we assume that the true TTF y is drawn from a Gaussian with unknown mean u and variance σ^2 and place a conjugate NIG prior on (u, σ^2):

$$y \mid u, \sigma^2 \sim \mathcal{N}(u, \sigma^2), \tag{14}$$

$$u \mid \sigma^2 \sim \mathcal{N}\left(\gamma, \sigma^2/v\right), \tag{15}$$

$$\sigma^2 \sim \text{Inv-Gamma}(\alpha, \beta) \tag{16}$$

Thus the tuple $(\gamma, v, \alpha, \beta)$ defines a distribution *over* possible Gaussian regressors rather than a single point estimate.

Prediction and uncertainty. The predictive mean of y is the posterior expectation of u,

$$\hat{y} = \mathbb{E}[u] = \gamma \tag{17}$$

At the same time, the NIG hyperparameters allow us to decompose the predictive variance into aleatoric and epistemic components:

$$\sigma^2_{\mathrm{alea}} = \mathbb{E}[\sigma^2] = \frac{\beta}{\alpha - 1} \qquad \sigma^2_{\mathrm{epi}} = \mathrm{Var}[u] = \frac{\beta}{v(\alpha - 1)} \tag{18}$$

Here σ^2_{alea} captures irreducible sensor and environmental noise, while σ^2_{epi} reflects model uncertainty due to limited or shifted data.

3.4 Training objective

For each window, the evidential head outputs $\Omega_t = (\gamma_t, \nu_t, \alpha_t, \beta_t)$ for the TTF target y_t, and the classification head predicts a probability p_t for $z_t \in \{0,1\}$. We collect all supervised negative log-likelihood terms in

$$\mathscr{L}^{\mathrm{NLL}}(\theta) = -\sum_t \log p(y_t, z_t \mid \mathbf{x}_{1:T}; \theta), \tag{19}$$

and use an evidence regularizer

$$\mathscr{L}^{\mathrm{R}}(\theta) = \sum_t |y_t - \gamma_t| \, (\nu_t + \alpha_t). \tag{20}$$

The overall uncertainty-aware loss in Fig. 1 is

$$\mathscr{L}(\theta) = \mathscr{L}^{\mathrm{NLL}}(\theta) + \lambda \, \mathscr{L}^{\mathrm{R}}(\theta). \tag{21}$$

4 Experiments

We now present the setup of our extensive experiments, followed by the details of the results of different evaluations.

4.1 Experimental setup

4.1.1 Datasets and label distribution

We use two synthetic datasets shared by NIST, both generated from compartment-fire simulations for flashover analysis. The first dataset corresponds to a three-compartment configuration[1] (ignition room–corridor–adjacent room). It contains temperature time series sampled every 20s from about 1000 numerical experiments (roughly 8 million time steps in total, with runs lasting up to 8400s). The second dataset corresponds to flashover in a six-compartment single-story dwelling[2], with temperature sequences sampled every 15s under multiple combinations of ignition locations and ventilation conditions, yielding 5041 simulated events.

For each dataset, we adopt 5-fold cross-validation. Table 1 summarizes the training and testing label distributions on both datasets. The data are highly imbalanced: in the three-compartment case the ratio of negative to positive samples is roughly 34:1, and in the six-compartment case it is about 69:1 in the training splits.

[1]The full dataset including the layout of the three-compartment can be found at `https://data.nist.gov/od/id/mds2-2258`.

[2]The full dataset including the layout of the six-compartment can be found at `https://data.nist.gov/od/id/mds2-2479`.

Table 1: Label distribution for three-compartment and six-compartment datasets across 5-fold cross-validation.

	Three-compartments				Six-compartments			
	Training		Testing		Training		Testing	
Fold	Pos	Neg	Pos	Neg	Pos	Neg	Positive	Negative
1	3,135	105,434	801	801	28,847	1,989,699	7,216	7,216
2	3,146	104,215	790	790	29,037	1,990,393	7,026	7,026
3	3,101	106,217	835	835	28,828	1,989,687	7,235	7,235
4	3,245	97,560	691	691	28,785	1,989,683	7,278	7,278
5	3,117	104,850	819	819	28,755	1,990,046	7,308	7,308

4.1.2 Baselines

We compare our method against a set of recurrent and attention-based sequence models widely used for time-series forecasting:

- **RNN**: a standard vanilla recurrent neural network that captures short-term dynamics via a simple recurrent state, used as a lightweight baseline.
- **GRU**: a gated recurrent unit with reset and update gates to mitigate vanishing gradients, suitable for modeling medium-range temporal dependencies.
- **LSTM**: a long short-term memory network with input/forget/output gates to capture long-term dependencies in a more robust manner.
- **Bi-RNN**: a bidirectional RNN that processes the sequence in both forward and backward directions to provide symmetric context representations.
- **Bi-GRU**: a bidirectional GRU that exploits past and future context simultaneously to enhance temporal feature extraction.
- **Bi-LSTM**: a bidirectional LSTM that learns long-range dependencies from both directions of the sequence.
- **P-Flash** [22]: a Bi-LSTM augmented with an attention mechanism that adaptively focuses on informative time steps and sensor channels.
- **P-Flashv2** [23]: a Bi-GRU equipped with an attention layer, aiming to highlight important temporal segments with a relatively small parameter budget.

We further include a CNN-LSTM baseline (**CNN-LSTM**) [14] and a recent graph-augmented flashover-specific model (**gFlashNet**) [10] as stronger task-oriented competitors.

4.1.3 Implementation Details

We implement all models in PyTorch. Unless otherwise stated, the hidden dimension is set to 128, the batch size to 64, and the initial learning rate to 1×10^{-4}. Parameters are optimized with Adam and trained for 10 epochs for each fold.

4.1.4 Evaluation metrics

In each fold, we construct windowed sequences with a history length of $k = 5$ and label them as positive if flashover occurs within the decision horizon, and negative otherwise. Following prior work, we truncate sensor readings to mimic sensor failure under high temperatures: the cut-off temperature is set to 200°C for the three-compartment dataset and 250°C for the multi-compartment dataset. As evaluation metrics, we report mean absolute error (MAE), accuracy (Acc), precision (P), recall (R), and F1-score. To ensure comprehensive validation, we keep the numbers of positive and negative samples balanced in the test set of each fold.

Table 2: Performance comparisons on flashover prediction. The best performance is in bold font.

	Three-compartments					Six-compartments				
Baselines	MAE ↓	Acc ↑	P ↑	R ↑	F1 ↑	MAE ↓	Acc ↑	P ↑	R ↑	F1 ↑
GRU	23.43	0.7661	**1.0000**	0.5321	0.6946	31.92	0.8704	0.9966	0.7433	0.8515
LSTM	21.16	0.8073	**1.0000**	0.6147	0.7614	35.68	0.8607	**0.9996**	0.7238	0.8386
RNN	24.12	0.7477	**1.0000**	0.4954	0.6626	33.37	0.8784	0.9965	0.7594	0.8614
Bi-GRU	18.69	0.8382	**1.0000**	0.6765	0.8070	30.02	0.8909	0.9963	0.7848	0.8780
Bi-LSTM	18.74	0.8392	**1.0000**	0.6765	0.8070	33.99	0.8758	0.9963	0.7565	0.8768
Bi-RNN	19.14	0.7679	**1.0000**	0.6596	0.7949	29.86	0.8958	0.9960	0.7948	0.8841
P-Flash	15.29	0.8716	**1.0000**	0.7413	0.8526	27.91	0.9088	0.9953	0.7961	0.8841
P-Flashv2	19.57	0.8486	**1.0000**	0.6972	0.8216	26.93	0.8992	0.9963	0.7873	0.8796
CNN-LSTM	19.8	0.8661	**1.0000**	0.7523	0.8586	31.78	0.8984	0.9957	0.8004	0.8874
gFlashNet	17.46	0.8575	**1.0000**	0.6750	0.8060	43.13	0.8698	0.9956	0.7426	0.8508
EviFlash(Ours)	**12.8**	**0.9135**	**1.0000**	**0.8270**	**0.9053**	**22.67**	**0.9285**	0.9964	**0.8711**	**0.9295**

4.2 Overall Performances

Table 2 summarizes the performance of all methods on the three-compartment and six-compartment datasets. On the three-compartment dataset, our model reduces MAE from 15.29 (Bi-LSTM-Attention) to 12.80 (about 16%) and increases F1 from 0.8586 (CNN-LSTM) to 0.9053; on the more complex six-compartment dataset, MAE drops from 26.93 (Bi-GRU-Attention) to 22.67 (also about 16%) and F1 rises from 0.8874 (CNN-LSTM) to 0.9295. Across baselines, bidirectional and attention-based variants (Bi-LSTM/GRU with attention, CNN-LSTM) consistently outperform their unidirectional counterparts, while our Mamba-based forecaster further improves both regression and classification metrics.

Because precision is almost saturated (close to 1.0) due to class imbalance, recall and F1 are more indicative of operational performance, where missed flashover alarms are far more critical than

false positives. Our model achieves the highest recall on both datasets (0.8270 and 0.8711), explaining the large F1 margins and suggesting earlier, more reliable warnings without sacrificing accuracy. Finally, gFlashNet, which is tailored to flashover prediction, is clearly outperformed—particularly on the six-compartment dataset—highlighting the benefit of combining a strong sequence backbone with explicit uncertainty handling.

4.3 Ablation studies

To quantify the contribution of our main design choices, we conduct ablation experiments on both datasets. Table 3 reports results for: (i) the full model (*Ours*); (ii) *w/o Mamba*, where the Mamba backbone is replaced by a standard recurrent encoder while keeping the uncertainty heads; and (iii) *w/o Uncertainty*, where the Mamba encoder is retained but trained as a deterministic regressor/classifier without aleatoric–epistemic modeling.

On the three-compartment dataset, removing the Mamba backbone increases MAE from 12.8 to 14.98 (about 17%) and reduces F1 from 0.9035 to 0.8845; on the six-compartment dataset, MAE rises from 22.67 to 24.39 (roughly 7.6%) and F1 drops from 0.9295 to 0.8998. This confirms that the selective state space structure brings consistent gains over conventional recurrent encoders.

Eliminating the uncertainty components (*w/o Uncertainty*) also degrades performance: MAE increases to 13.95 and 23.92 on the three- and six-compartment datasets, and F1 falls to 0.8894 and 0.9003. Although this deterministic variant still outperforms most baselines, it remains worse than the full model, indicating that explicitly modeling aleatoric and epistemic uncertainty not only improves calibration but also yields measurable gains in point accuracy and classification quality. Overall, the ablations show that both the Mamba backbone and the uncertainty-aware heads are necessary to achieve the best flashover prediction performance.

Table 3: Ablation Experiments

	Three-compartments		Six-compartments	
Methods	MAE ↓	F1 ↑	MAE ↓	F1 ↑
EviFlash(Ours)	**12.8**	**0.9035**	**22.67**	**0.9295**
w/o Mamba	14.98	0.8845	24.39	0.8998
w/o Uncertainty	13.95	0.8894	23.92	0.9003

4.4 Analysis of sensor cut-off temperatures

We mimic high-temperature sensor failure by truncating heat-detector readings above a cut-off temperature T_{cut}. Fig. 2 reports the F1-score of different methods as T_{cut} varies on the three-compartment and six-compartment datasets, respectively.

On the three-compartment dataset (left subfigure in Fig. 2), all methods achieve relatively high F1 when T_{cut} is close to the nominal failure point (e.g., 165°C), and performance degrades as T_{cut} is lowered, i.e., as the sensor fails earlier and more peak information is lost. The degradation patterns, however, are quite different: recurrent baselines and their attention variants experience a sharp

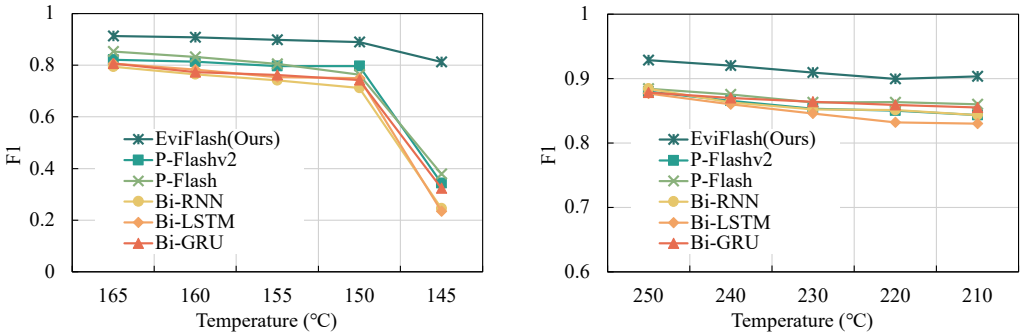

Fig. 2. Heat detector fails at different temperatures respectively in three-compartments and six-compartments.

collapse when $T_{\text{cut}} \leq 150^{\circ}\text{C}$, with F1 dropping below 0.4, whereas our model degrades much more gracefully and still maintains an F1 above 0.8 at the most severe truncation. This suggests that the proposed selective state space forecaster can exploit temporal and cross-sensor redundancy to compensate for missing high-temperature segments.

On the six-compartment dataset (right subfigure in Fig. 2), the impact of truncation is milder but the trend remains consistent. As T_{cut} decreases from 250°C to 210°C, all methods show a gradual reduction in F1, yet our approach remains several points above the strongest baselines across all cut-off settings. These results indicate that, even under realistic scenarios where heat detectors saturate or fail at different temperatures, our uncertainty-aware Mamba model preserves higher detection quality and supports more reliable flashover warning in the presence of sensor saturation and failure.

4.5 Transferability analysis

In real deployments, the compartment geometry, ventilation pattern, and detector placement often differ substantially from those in the training data, leading to domain shift in the statistics and temporal patterns of sensor readings. To assess how well different models cope with such shift, we perform a transferability study: each model is trained on one configuration and then evaluated directly on the test split of the other configuration without any fine-tuning.

Figure 3 shows results when models are trained on the three-compartment dataset and evaluated on the six-compartment dataset. Most bidirectional recurrent baselines, with or without attention, degrade to accuracies close to 0.50, i.e., only slightly above random guessing under severe class imbalance. Our method, by contrast, reaches an accuracy of about 0.63, outperforming the best baseline by a clear margin. This indicates that the proposed selective state space forecaster leverages temporal and cross-sensor regularities that remain informative even as room layout and ventilation conditions change.

A similar pattern appears in the reverse direction (right subfigure Fig. 3), where models are trained on the six-compartment dataset and tested on the three-compartment dataset. Conventional bidirectional RNN/GRU/LSTM variants cluster around 0.50–0.58 accuracy, whereas our model at-

tains roughly 0.64. The consistent gains in both transfer directions suggest that the Mamba backbone, together with the uncertainty-aware heads, yields more robust representations under geometric and sensor-layout shifts, leading to stronger cross-configuration generalization and practical transferability.

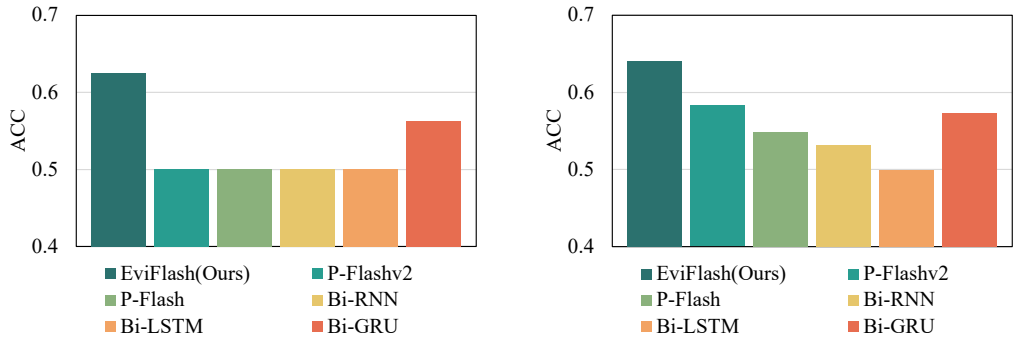

Fig. 3. Transferability analysis respectively for three-compartments to six-compartments (left subfigure) and for six-compartments to three-compartments (right subfigure).

5 Conclusion

We proposed EviFlash, an uncertainty-aware flashover prediction framework built on a selective state space model. By casting multi-sensor flashover forecasting as probabilistic sequence modeling, EviFlash jointly predicts time-to-flashover and flashover events while decomposing predictive uncertainty into aleatoric and epistemic parts. This design enables the model not only to produce point estimates, but also to provide more informative confidence information for risk-aware decision making in early warning scenarios. Experiments on two NIST compartment-fire datasets demonstrate that EviFlash achieves consistent gains over recurrent and attention-based baselines in both MAE and F1. In particular, the proposed framework shows stronger performance in detecting imminent flashover and maintaining stable regression accuracy under challenging conditions. Ablation results further confirm that both the selective state space backbone and the uncertainty modeling contribute to the final improvement. Moreover, the proposed framework exhibits graceful degradation under sensor cut-off and strong cross-layout transferability, indicating improved robustness when sensor readings are incomplete or the compartment geometry changes. Overall, these findings suggest that uncertainty-aware sequence modeling is a promising direction for early flashover warning, especially for real-world firefighting applications that require both accuracy and reliability.

Acknowledgments

This work was supported in part by the National Key R&D Program of China (2024YFC3016500, 2024YFC3016503) and the Central Public-interest Scientific Institution Basal Research Fund (20248803Z, 20238801Z, Z20248810).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] Graham T, Makhviladze GM, Roberts JP. On the theory of flashover development. Fire Safety Journal. 1995;25(3):229-59.

[2] Moore A. Apology call after nightclub fire deaths verdict; 2024. BBC News, [EB/OL]. Available from: https://www.bbc.com/news/articles/cn0rvgg7dlzo.

[3] Ministry of Emergency Management of the People's Republic of China. 165 people sacrificed in rescue missions over the past five years; 2023. Official Press Release, [Online]. Available from: https://www.mem.gov.cn/xw/xwfbh/2023n11y7rxwfbh/mtbd_4262/202311/t20231107_467922.shtml.

[4] Thomas PH. The growth of fire—mathematical interpretation. Fire Research Abstracts and Reviews. 1967;9(1):1-8.

[5] Overholt KJ, Ezekoye OA. Characterizing heat release rates using an inverse fire modeling technique. Fire Technology. 2012;48(4):893-909. Received 29 April 2011 / Accepted 26 December 2011.

[6] Overholt KJ, Ezekoye OA. Inverse heat release rate estimation using a two-zone fire model and genetic algorithm. Fire Safety Journal. 2012;51:53-62.

[7] Haworth DC. Progress in probability density function methods for turbulent reacting flows. Progress in Energy and Combustion Science. 2010 January;36(2):168-259.

[8] Peacock RD, Forney GP, Reneke PA. CFAST – Consolidated Model of Fire Growth and Smoke Transport (Version 6): Technical Reference Guide. Gaithersburg, MD: National Institute of Standards and Technology; 2013. 1026r1. Available from: https://doi.org/10.6028/NIST.SP.1026r1.

[9] Song Y, et al. Machine Learning Based Flashover Prediction Models Using Synthetic Data and Fire Images. Fire Technology. 2025. Vol. 61, no. 1, Feb. 2025. Available from: https://link.springer.com/article/10.1007/s10694-024-01440-7.

[10] Fan L, Fang H, Liang T, Tam WC, Zhang Q. A cost-effective data-driven approach to flashover prediction across diverse residential layouts for enhanced firefighters situational awareness. Journal of Building Engineering. 2025;100:111728.

[11] Mozaffari MH, Li Y, Hooshyaripour N, Ko YJ. Vision-Based Prediction of Flashover Using Transformers and Convolutional Long Short-Term Memory Model. Electronics. 2024 December;13(23):4776.

[12] Smirnov NN, Betelin VB, Nikitin VF, Stamov L. Accumulation of errors in numerical simulations of chemically reacting gas dynamics. Acta Astronautica. 2015 December;117:338-55.

[13] Chung WT, Mishra AA, Perakis N, Ihme M. Data-assisted combustion simulations with dynamic submodel assignment using random forests. Combustion and Flame. 2021;227:172-85.

[14] Lu X, He M, Wang Z, Hu H, Ji J, Zhu J. Prediction of flashover time in a compartment fire by CNN-LSTM based deep neural network considering wearable data collection equipment. Journal of Building Engineering. 2024;97:110719.

[15] Joshi AS, William GJ, Joshi SS. Uncertainty-Aware Post-Detection Framework for Enhanced Fire and Smoke Detection in Compact Deep Learning Models. arXiv preprint arXiv:251010108. 2025. October 2025.

[16] Pitas K, Arbel J. Something for (almost) nothing: Improving deep ensemble calibration using unlabeled data. arXiv preprint arXiv:231002885. 2023.

[17] Angelopoulos AN, Bates S, et al. Conformal prediction: A gentle introduction. Foundations and Trends® in Machine Learning. 2023;16(4):494-591.

[18] Gibbs I, Candes E. Adaptive conformal inference under distribution shift. Advances in Neural Information Processing Systems. 2021;34:1660-72.

[19] Fellaji M, Pennerath F, Conan-Guez B, Couceiro M. On the Calibration of Epistemic Uncertainty: Principles, Paradoxes and Conflictual Loss. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer; 2024. p. 160-76.

[20] Gu A, Dao T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. In: First Conference on Language Modeling; 2024. Available from: https://openreview.net/forum?id=tEYskw1VY2.

[21] Tai W, Chen B, Zhou F, Zhong T, Trajcevski G, Wang Y, et al. TrustGeo: Uncertainty-aware dynamic graph learning for trustworthy IP geolocation. In: Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining; 2023. p. 4862-71.

[22] Fu EY, Tam WC, Wang J, Peacock R, Reneke PA, Ngai G, et al. Predicting flashover occurrence using surrogate temperature data. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35; 2021. p. 14785-94.

[23] Tam WC, Fu EY, Li J, Peacock R, Reneke P, Ngai G, et al. Real-time flashover prediction model for multi-compartment building structures using attention based recurrent neural networks. Expert Systems with Applications. 2023;223:119899.