This is the Pre-Published Version.
Tam et al (2022) *A Spatial Temporal Graph Neural Network Model for Predicting Flashover in Arbitrary Building Floorplans*, **Engineering Applications of Artificial Intelligence**, 115, 105258. https://doi.org/10.1016/j.engappai.2022.105258

A Spatial Temporal Graph Neural Network Model for Predicting Flashover in Arbitrary Building Floorplans

Wai Cheong Tam[1,†,*] , Eugene Yujun Fu[2,†], Jiajia Li[1], Xinyan Huang[3], Jian Chen[4], Michael Xuelin Huang[5]

[1]National Institute of Standards and Technology, Gaithersburg, Maryland, USA
waicheong.tam@nist.gov

[2]Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
eugene.fu@polyu.edu.hk

[3]Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
xy.huang@polyu.edu.hk

[4]Center for Offshore Engineering and Safety Technology, China University of Petroleum (East China), Qingdao, China
chenjian@upc.edu.cn

[5]Google, Inc.
mxhuang@google.com

ABSTRACT

Rapid fire progression, such as flashover, has been one of the leading causes for firefighter deaths and injuries in residential building environments. Due to long computational time of and the required prior knowledge about the fire scene, existing models cannot be used to predict the potential occurrence of flashover in practical firefighting applications. In this paper, a scene-agnostic model (FlashNet) is proposed to predict flashover based on limited heat detector temperature information up to 150 °C. FlashNet utilizes spatial temporal graph convolutional neural networks to effectively learn features from the limited temperature information and to tackle building structure variations. The proposed model is benchmarked against five different state-of-the-art flashover prediction models. Results show that FlashNet outperforms the existing flashover prediction models and it can reliably predict flashover 30 seconds preceding its occurrence with an overall accuracy of about 92.1 %. Ablation study is carried out to examine the effectiveness of different key model components and geometric average adjacency matrix. The research outcomes from this study are expected to enhance firefighters' situational awareness in the fire scene, protecting them from hazardous fire environments and to pave the way for the development of data-driven prediction systems.

Keywords: Machine learning; compartment fires; synthetic fire data; smart firefighting; intelligent system

1 INTRODUCTION

In the United States, firefighter deaths and injuries are the mounting concerns. Approximately 800 firefighters have been killed (Fahy et al. 2020) and more than 320,000 injured (Campbell et al. 2019) from 2008 to 2018. Many of these injuries ended their careers. The National Fire Protection Association (Fahy et al. 2020) suggested that approximately 58 % of fireground deaths occurred in residential areas. Studies (Fahy et al. 2020 and Campbell et al. 2019) also revealed that rapid fire progression, such as flashover, was responsible for 13 % of firefighter deaths and numerous injuries. A *flashover* is the near-simultaneous ignition of most of the directly exposed combustible material in an enclosed area (Peacock et al. 1999), and

[†] Joint first authors.

its occurrence is often characterized by a smoke layer temperature of about 600 °C (Reneke 2013, Hurley et al, 2015). However, firefighters have not had any tools to predict flashovers and rely solely on their experience to recognize flashover indicators (Hamins et al. 2015), such as rollover near ceiling and/or high heat, to avoid this life-threatening event. Yet, since the fire conditions change rapidly and the transition of flashover usually happens within seconds (Garrity and Yusuf 2021), it is rather difficult to identify the imminent occurrence of flashover while inside a building fire and hardly possible on the outside. With that, if the fire fighters cannot recognize impending flashover their lives are at tremendous risk. It is believed that with the advancement of smart interconnected fire protection devices and systems (Reneke 2013), a data-driven model can be developed to predict the potential occurrence of flashover based on typical thermal sensor signals, such as from heat detectors.

Over the past twenty years, many attempts have been made to develop real-time prediction models. Existing prediction models include empirical correlations (Babrauskas 1980, McCaffrey et al. 1981, Mitler and Steckler 1995, Kim and Lilley 2002), inversed modelling techniques (Overholt and Ezekoye 2012, Price 2014), and sensor-steered computational fluid dynamics (CFD) approaches (Koo et al. 2010, Jahn 2017). However, the empirical models, such as (Babrauskas 1980, Mitler and Steckler 1995, Kim and Lilley 2002), are limited to single-compartment applications and are not suitable for multi-compartment structures. The sensor-steered CFD approaches (Koo et al. 2010, Jahn 2017) are more generic for structure variance. They provide better prediction capabilities and a higher accuracy. However, these models require expensive computation and rely on high-performance computing machines. For instance, one simulation time step takes more than five minutes to compute (Jahn 2017). Therefore, CFD approaches fail to meet the need for real-time firefighting applications.

Other limitations associated with the above-mentioned models are that they rely on both the continuous temperature signals from thermocouples and the prior knowledge about the fire scene. The *fire scene* information includes building structures, fire location, fire size, fire growth rate, and vent opening conditions of doors and windows. Residential fire protection devices, such as heat detectors, will stop functioning at elevated temperature (i.e., 150 °C) (NFPA 2002), and the information about the fire locations and the vent conditions are generally unknown. Since these realistic conditions (i.e., sensor temperature limit and the effect of arbitrarily fire location and vent opening conditions) have not been considered in the model development process, there is a knowledge gap between existing theoretical studies and the existing solutions to practical firefighting applications. This present study thus aims to fill this gap to overcome both the challenges of real-time computation and the complexity of realistic conditions.

Machine learning (ML) has made breakthroughs in various scientific and engineering problems, including fault detection (Nasiri et al. 2017), cyber-attack prevention (Rao and Frtunikj 2018), object detection for self-driving vehicles (Brundage et al. 2018), human activity recognition (Xiao et. al 2021a), offline learning (Swazinna et. al, 2021), and motor imagery EEG subject classification (Arunabha Roy 2022), where real-time predictions and reliable accuracy are needed. Efforts of ML-driven fire forecast have also been made in the fire research community. Hodges et al. (2019) used a neural network model to estimate smoke spread in a T-shape structure. Tam et al. (2020) developed a fire data generator to facilitate the use of ML for flashover prediction in multi-compartment structures. Dexters et al. (2021) used a classic ML model to determine the relationships for different fire parameters. Garrity and Yusuf (2021) utilized a simple feed-forward neural network model to forecast the temperature rise in a standard testing enclosure. Wu et al. (2021) and Zhang et al (2022) used deep learning models to predict local temperature in built environments. However, the dataset being considered in these studies were generally limited and simplifications were made. Wang et al. (2021) investigated ML paradigms to predict flashover based on limited sensor temperature up to 150 °C. In Fu et al. (2021), the ML-based flashover prediction model was expanded to account for a wider range of fire scenarios. The model was also validated against real fire experiments and prediction accuracy of ~ 82 % with a lead time of 30 seconds preceding flashover events was reported. In the same study, it was demonstrated that the required computational time for each prediction took less than one second. Compared with the over five-minute simulation in Koo et al. (2010) and Jahn (2017), the improvement in computation efficiency is substantial and the use of machine learning approach is promising.

Practical conditions are needed to be considered for firefighting applications and the main limitation of existing prediction models such as Wang et al. (2021) and Fu et al. (2021) lies in their generalizability across building structures. Their model architectures do not support a variable number of input data and floorplans. For example, a model trained for one building structure with four channels of temperature data does not generalize to another with eight channels. This model limitation imposes the need of prior knowledge about the exact floorplan of the building structure. Yet, this kind of information is usually unknown in practical firefighting. To account for fire scene diversity, Flashover prediction neural Network (FlashNet) is proposed as a scene-agnostic flashover prediction model. It formulates the building structures with different floorplans as graphs, whose nodes represent the sensor inputs in each building compartment. FlashNet uses graph neural networks (GNNs) to learn the signal dependency among the compartments and predicts flashover from the sensor signals and their interrelationships. Conceivably, the use of GNNs allows FlashNet to implicitly understand different fire scenarios, such as building structure, fire locations, and vent opening conditions.

The contributions of this work are as follows: 1) to the best of the authors' knowledge, this is the first study to apply GNNs to achieve a scene-agnostic flashover prediction model, 2) to use a comprehensive synthetic dataset for the development of a machine learning based flashover prediction model in multi-compartment structures with a wide range of fire and vent opening conditions, 3) to propose the formulation of a geometric average adjacency matrix which opens a new door to handle fire data from different building structures, 4) to propose a novel GNN model that effectively captures the temporal information about the limited heat detector temperature signals and the spatial relationships between different compartments for flashover predictions, and 5) results show that FlashNet outperforms the existing state-of-the-art flashover prediction models and achieves reliable model performance for flashover prediction 30 seconds preceding the event without any prior knowledge about the fire scene. It is believed that the outputs from this work will provide a step forward to facilitate smart firefighting (Hamins et. al 2015) and help reduce fire fighter deaths and injuries.

The next section will discuss the data collection process. Section 3 presents the model formulation and Section 4 discusses the results.

2 TRAINING DATA SYNTHESIS FROM SIMULATION

Training a well-performing ML model requires an adequate dataset. However, the acquisition of sufficient real flashover fire test data for diversity of home structures is expensive, if even possible. Building a model for a three-compartment building structure with simplified fire and vent opening conditions would require approximately 1000 full-scale simulated experiments in a recent study (Tam et al. 2020). Considering the diversity of possible fire scenes, the required number of experiments increases exponentially with the complexity of buildings and becomes impossible to achieve. As such, this study applies the learning-by-synthesis approach (Sugano et al. 2014) and trains the flashover model using massive simulated fire data.

This section first describes the fire data synthesis engine and then the set of variables in fire data synthesis, including home structures, compartment configurations, fire conditions, vent opening conditions, and sensor temperature profiles.

2.1 Fire Data Synthesis Engine

Training data for FlashNet is synthesized by a zone model, namely CFAST (Peacock et. al 2021). CFAST is a two-zone fire model that predicts the thermal environment in a compartmented structure. The model is validated against more than 15 sets of full-scale experiments (Peacock et al. 2015) with peak heat release rate (HRR), compartment aspect ratio (i.e., compartment length versus ceiling height), and global equivalence ratio, ranging from approximately 50 kW to 15,700 kW, 0.4 to 4.9, and roughly 0 to a value larger than 1 for a wide range of ventilation factors, respectively. In terms of accuracy, the average discrepancy of CFAST predictions of upper layer gas temperature against the experimental measurements is within 6 %. Appendix A provides a validation case for a 6-compartment home, and comparison shows

that the magnitudes and trends of the temperature profiles match well with the experimental data and the maximum absolute root mean squared error is about 30 °C.

CData (Reneke et al. 2019) is the input file generator for CFAST. It systematically generates CFAST input files, saving users from the time-consuming manual configuration work. The input file generation depends on a set of user-specified simulation parameters, including building layouts, surface materials, fire conditions, ventilation configurations, location of detector(s), and output intervals. CData also accounts for seven different probability density functions (i.e., uniform, normal, log-normal, etc.) to ensure the diversity of input file generation. Additionally, user-defined functions can also be used to formulate inputs for desired fire environments.

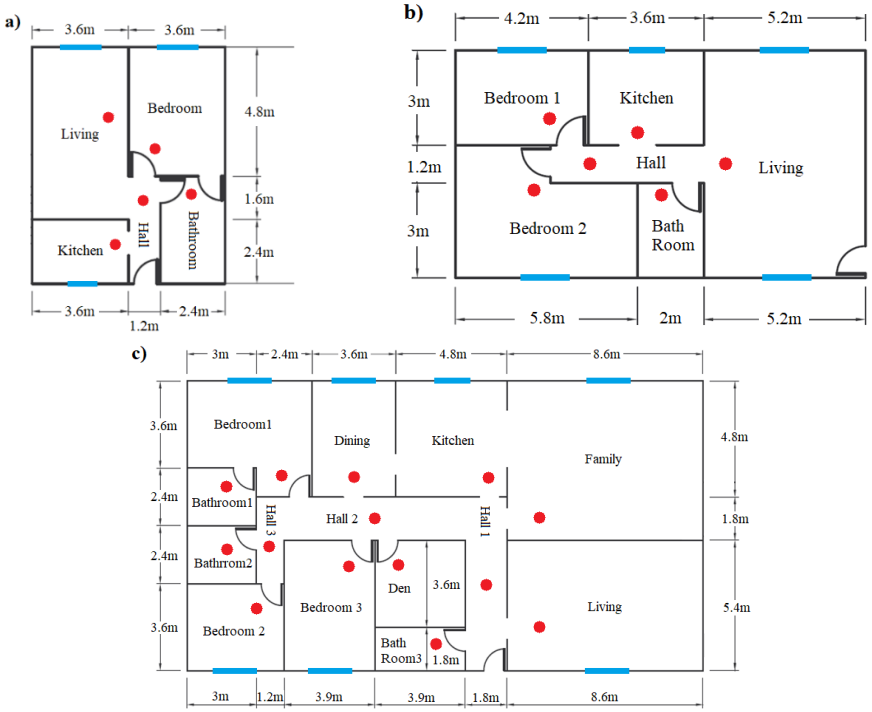
In this study, the data generation process was done on a Fire Research Division Linux Cluster using approximately 250 nodes, and it took about two weeks. A summary spreadsheet of all parameters for each case was generated. It allows for data inspection, such as removing duplicate cases and examining data behavior of each simulation parameter. This information is crucial to ensure the quality and range of the fire data.

2.2 Variables of Fire Data Synthesis

The parameters of fire data synthesis include home structures, compartment configurations, fire conditions, vent opening conditions, and sensor profiles. The following sections will discuss each of these factors.

2.2.1 Home structures

Seventeen typical single-floor home structures are selected from (Persily et al. 2006), which defines 209 dwellings to represent approximately 80 % of U.S. housing layouts. These 17 structures can be categorized into three types of residential buildings: 1) apartment homes, 2) attached homes, and 3) detached homes . The overall floor area ranges from 65 m^2 to 275 m^2 with three to fourteen compartments. Figures 1 shows examples of an apartment home (a), an attached home (b), and a detached home (c). The complete list of the 17 homes is given in Appendix C. The hallway in Fig. 1c is divided into three sections for data generation. Table 1 describes the floor area and interior details of the 17 homes.



Figures 1. Examples of the building layout for a) a five-compartment apartment home (#2), b) a six-compartment attached home (#9), and c) a fourteen-compartment detached home (#17).

To properly train and evaluate the scene-agnostic model, 17 representative home structures were carefully selected to include buildings with identical numbers of compartments but different layouts (i.e.,

Home #4 and Home #6 in Table 1) and buildings with different numbers of compartments. The exploration of the extensive dwelling structures requires different research efforts and is beyond the scope of this paper.

Table 1. Housing details for the seventeen selected homes for data generation.

Home #	Type*	Total Compartment	Floor Area (m^2)	Living Room	Kit-chen	Hall‡	Bed-Room	Bath-Room‡	Dining Room	Family Room	Den	Fire Cases
1	APT	3	65	1	1	0	0	1	0	0	0	2000
2	APT	5	65	1	1	1	1	1	0	0	0	3000
3	APT	6	65	1	1	2	1	1	0	0	0	3000
4	APT	7	65	1	1	2	2	1	0	0	0	4000
5	APT	6	65	1	1	1	2	1	0	0	0	4000
6	APT	7	65	1	1	1	2	2	0	0	0	4000
7	APT	8	65	1	1	1	2	2	0	0	1	5000
8	AH	5	95	1	1	1	1	1	0	0	0	3000
9	AH	6	95	1	1	1	2	1	0	0	0	4000
10	AH	7	95	1	1	1	2	1	1	0	0	5000
11	DH	7	107	1	1	1	3	1	0	0	0	5000
12	DH	7	107	1	1	1	2	1	1	0	0	5000
13	DH	8	107	1	1	1	3	1	1	0	0	6000
14	APT	8	130	1	1	2	2	2	0	0	0	4000
15	APT	8	142	1	1	2	2	1	0	0	1	5000
16	DH	13	275	1	1	3	4	2	1	1	0	8000
17	DH	14	275	1	1	3	3	3	1	1	1	8000

* Note that apartment home, attached home, and detached home are denoted as APT, AH, and DH, respectively. ‡ No fire in these compartments. In total, there are 78,000 cases for the seventeen different structures.

2.2.2 Compartment configurations

The compartment configurations are similar for the selected home structures. The overall ceiling height is 2.4 m. The height and width for doors and doorways are 2.05 m and 0.9 m, respectively. The horizontal location of each window (shown as blue lines) is centred in a compartment wall as seen in Figs. 1. The width, height, and vertical locations of the windows are 1.8 m x 1.4 m x 0.6 m for the living room, 1.4 m x 0.85 m x 1.2 m for the kitchen, 1.4 m x 1.4 m x 0.6 m for the dining room, 1.8 m x 1.4 m x 0.6 m for the family room, and 1.4 m x 1.4 m x 0.6 m for the bedrooms. The glazing material is consistent for all windows, taken as 3 mm single-pane float glass (Hurley et al. 2015). The thermal properties are shown in Table 2.

Every compartment contains one heat detector, which generally is about 0.02 m below the ceiling and near doors or doorways as shown in red dots in Figs. 1. The thermal response time index for the heat detector is taken as 35 $(m \cdot s)^{0.5}$. Temperature data obtained from the heat detectors are used for model development. Regarding the interior finish, the walls and the ceiling are constructed with gypsum wallboards, and the floor is built with concrete. The thermal properties of the materials are depicted in Table 2.

Table 2. Summary of thermal properties for building materials (Hurley et al. 2015, Reneke et al. 2019).

Materials	Conductivity W/(m-K)	Specific Heat kJ/(kg-K)	Density kg/m^3	Thickness m	Emissivity (-)
Glazing	0.80	0.80	2500	0.003	0.95
Gypsum	0.16	1.00	480	0.025	0.90
Concrete	1.60	0.75	2400	0.150	0.94

2.2.3 Fire conditions

In each numerical experiment, there is one fire initiated in one compartment except the hallways and bathrooms. The fire is located at the center of the compartment. Each compartment may have one of three furniture items: chairs, polyurethane foam mattresses, and cotton-based spring mattresses. Fire growth typically contains linear growth (smoldering fire), t-squared growth (flaming fire), peak, and decay, as shown in Fig.2. Table 3 illustrates the fire growth parameters of the three furniture items, including the

transition heat release rate (HRR) from smoldering to a flaming fire (Q_1), peak HRR (Q_{max}), time to transition (t_1), time to peak HRR (t_2), peak time ($t_3 - t_2$), and decay time ($t_4 - t_3$).

The peak HRR and time to peak HRR are obtained from (Reneke et al. 2019), and the fire growth rate is between 3.29E-4 kW/s^2 and 4.14E-2 kW/s^2, ranging from slow to fast fire growth rate. Transition HRR and time to transition HRR are selected based on examination of HRR data provided in (Cleary 2014). Since flashover usually occurs during the t-squared growth stage and peak stage, the exact value of the peak and decay times are not significant. One thousand different fire cases are assigned to each compartment, and the total number of fire cases for each building structure is shown in Table 1.

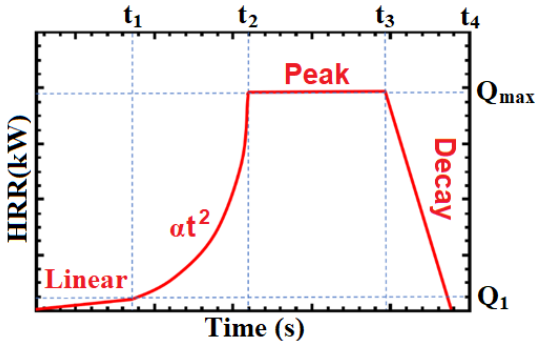


Figure 2. Heat release rate in four stages of fire growth for a typical t-squared fire.

Table 3. Heat release rate parameters for chair and mattresses (Reneke et al. 2019).

Items	Q_1 (kW)	Q_{max} (kW)	t_1 (s)	t_2 (s)	t_3-t_2 (s)	t_4-t_3 (s)
Chair	10 – 30	270 – 3500	150 – 1200	295 – 675	1000	100
Mattress (foam)	10 – 30	2275 – 4620	150 – 1200	305 – 435	1000	100
Mattress (cotton)	10 – 30	130 – 1670	150 – 1200	360 – 1240	1000	100

CData uses probability density functions (PDF) for samplings. The PDF is based on numerical experiments to include a wide range of fire growth behaviors and to provide sufficient cases that lead to flashover. Given the fact that a ML based model is intended to be developed for potential flashover prediction and since there will be a significant data imbalance associated with non-flashover related and flashover related data, a uniform PDF is used to sample peak HRR for the three different items, and a skew normal PDF is used to sample the calculated time to peak HRR. The motivation is that more flashover cases (i.e., more rapidly growing fires) can be included in the data set so that the ML based model will determine the important relationship about potential flashover conditions. For transition HRR, time to transition, peak time, and decay time, a uniform PDF is utilized.

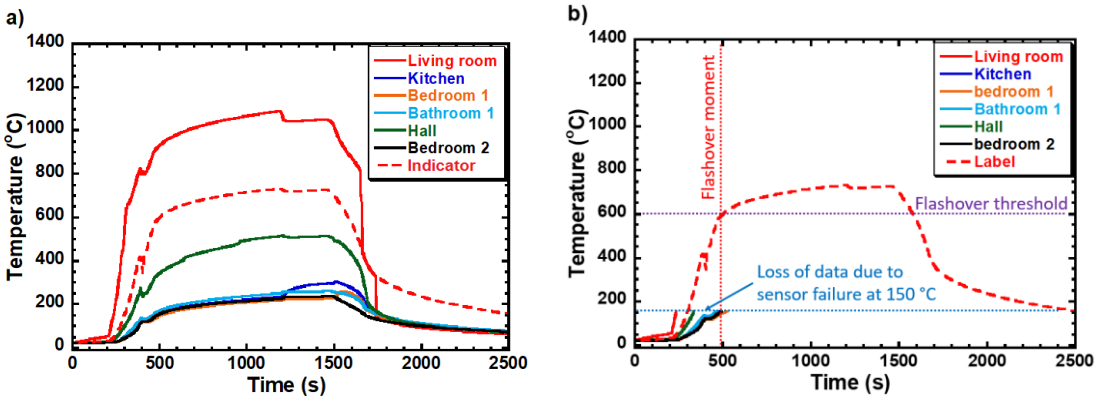
2.2.4 Vent opening conditions

There are three types of vents in this study: interior doors such as bedroom and bathroom doors, exterior doors such as front doors, and windows. For interior doors, the current setting assumes all interior doors to be opened from the beginning of all tests. This simplification is made to reduce the complexity of the data behavior; this constraint will be removed in future studies. For front doors, there is a 50 % chance that the door is open at the beginning of a test; providing fresh air to sustain fire growth. In terms of CData specifications, a uniform PDF is used to sample the opening of a door. For windows, a temperature-trigger setting is used to allow the windows to be arbitrarily opened when a temperature set-point is reached. Based on (Hurley et al. 2015), breakage of a single-pane float glass is experimentally observed at temperatures between 100 °C and 200 °C. An average value of 150 °C is used as the temperature set-point. In order to adopt the window breakage phenomenon, a target is placed at the top of a window. The direction of the target is normal to the window surface and the target thermal properties are taken to be that of a 3 mm single-pane float glass as shown in Table 1. It is worth noting that although the current fire and vent opening conditions are relatively simple, the generated data do cover a wide range of realistic fire scenarios. Also, given the fact that data is generated for 17 different buildings, the data size is substantial. For the numerical

setting mentioned above, there are a total of 78000 cases considered here. A complete table of parameters used for data collection is provided in Appendix B.

2.2.5 Sensor temperature profiles

Figure 3a shows six heat detector local temperature readings for a synthetic fire case with a medium growth fire (~ 0.014 kW/s^2) with high peak HRR (~ 3060 kW) occurring in the living room of an attached home as shown in Fig. 1b. The total simulation time for each numerical experiment is 3600 s, and the temperature output interval is 5 s. In this case, all interior and front doors are always open. Given sufficient oxygen, the fire continues to grow, and the detector temperature increases accordingly until the fire goes out. The temperature oscillation around 400 s is due to window breakage in the living room. It should be noted that the indicator (red dash line) is the upper layer gas temperature for the room of fire origin and is used to determine when the temperature threshold for the flashover condition is met.



Figures 3. Temperature profiles of heat detectors in different compartments a) without sensor limits and b) with a sensor limit of 150 °C for case #9 as shown in Fig.1b. Temperature profiles are only shown up to 2500 s.

A proper temperature criterion is carefully selected for flashover conditions. In (Peacock et al. 1999), the onset of flashover temperatures ranging from 450 °C to 771 °C. This wide range of values for the measurement criterion is due to the huge change of temperature during the transition to flashover. Yet, (Reneke 2013) observes that most of the values are in the 550 °C to 650 °C range. To be conservative and following the recommendations from (Hurley et al, 2015), an upper gas layer temperature of 600 °C is used as the threshold for the potential occurrence of flashover in this study. Flashover, of course, is also dependent on the presence of a sufficient quantity of oxygen.

The loss of a heat detector temperature signal is a realistic fire damage condition that adds another layer of complexity to the current problem. In actual fire scenarios, heat detectors are very unlikely to survive at elevated temperatures (Pomeroy 2010) approaching flashover conditions and will fail at temperatures well below flashover occurrence. According to NFPA 72 (NFPA 2002), heat sensing fire detectors are categorized into seven different classes with temperature ranging from low to ultra-high and the maximum operational temperature ranging from 29 °C to 302 °C. In this study, a temperature cut-off of 150 °C is adapted. As an illustration, Fig. 3b depicts the temperature profiles as shown in Fig. 3a with a temperature cut-off at 150 °C. With less available temperature information, the prediction of potential flashover occurrence becomes more difficult. Therefore, in addition to being able to determine different building layouts, the model also needs learning capabilities to correlate complex temperature information from other compartments as well as flashover conditions in the room of fire origin.

3 GENERIC FLASHOVER PREDICTION MODEL (FLASHNET)

3.1 Overall Model Structure

In this section, the overall structure of the generic flashover prediction model (FlashNet) is presented. As shown in Fig. 4, FlashNet is composed of two different modules. In the first module, the multivariate temperature inputs are transformed into graph representations and graph-structured data $\mathcal{G} = (X, A)$ are then formed where X is the node-attribute matrix and A is the adjacency matrix. For this study, the node-

attribute matrix contains temperature information from the heat detectors. Consider Fig.1b as an example, there are six compartments. The current numerical setting assumes one heat detector in one compartment, $X \in \mathbb{R}^{M \times N \times O}$ where M is the length of the temporal temperature sequence, N is the number of compartments, and O is the number of attributes. Since only temperature is used, O is equal to one and N is equal to six for the building structure shown in Fig.1b. For the adjacency matrix, $A \in \mathbb{R}^{N \times N}$ in which $a_{i,j}$ is non-zero if compartment i and compartment j are connected. In principle, $a_{i,j}$ indicates the strength of the connection. If there is no connection between the two compartments, then $a_{i,j}$ is zero. Additional discussion is provided in the following subsections to explain how the adjacency matrix is determined.

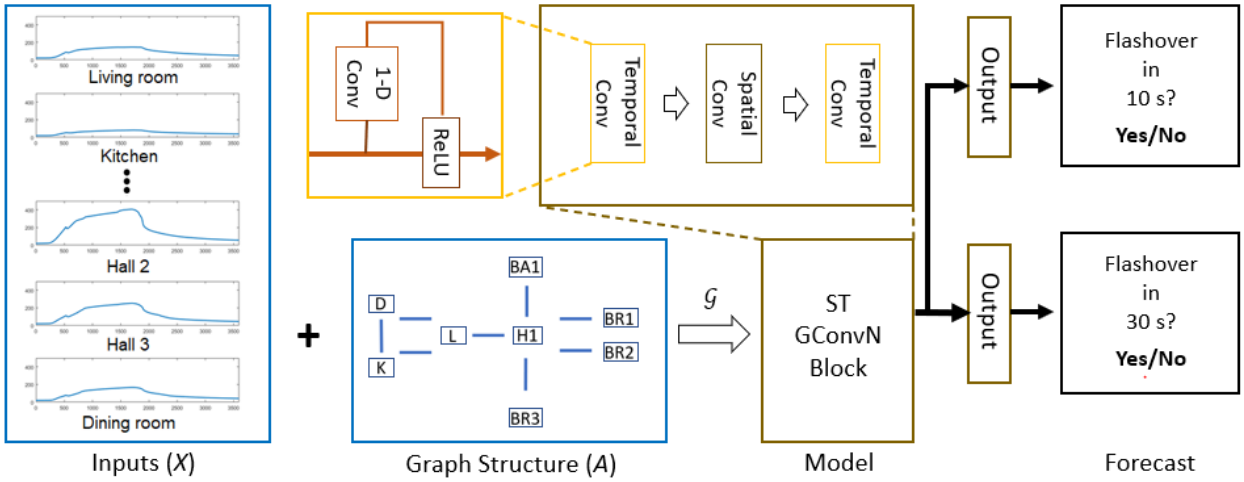


Figure 4. Model structure of FlashNet. The framework consists of the graph-structured data formulation module and the computational module.

In the second module, the graph data $\mathcal{G}$ serves as inputs to the spatial temporal graph convolutional networks (ST-GConvN). The ST-GConvN block is used to capture the spatial and the temporal dependencies from the multivariate temperature data and consists of two temporal convolution layers and one spatial graph convolution layer. For each temporal convolution layer, there is a 1-D convolution following by a rectified linear unit (ReLU), and this operation provides additional non-linearity to the temporal convolution layer to learn important features/information (i.e., how fast the temperature increase, how much time it takes to reach a certain temperature condition, etc.) in the time domain. Mathematically, the relationship between the inputs and the outputs in the fire temporal convolution layer is given as:

$$H = h^1(Z) = ReLU(Z^1 * W_t^1 + b_t^1) \tag{1}$$

where $Z^i \in \mathbb{R}^{M \times N \times C_i}$ is the input in a general form, $H \in \mathbb{R}^{(M-k_t+1) \times N \times C_o}$ is the output, $W_t^1 \in \mathbb{R}^{k_t \times C_i \times C_o}$ is the weights, and $b_t^1 \in \mathbb{R}^{C_o}$ is the biases for the first temporal convolution layer. h is the temporal convolution operation and C_i and C_o are the node numbers for inputs and outputs, respectively. The temporal convolution layer uses a k_t-width kernel to perform temporal convolution operation and encodes temporal dependencies from k_t neighbours of the input elements (i.e., $t-2$, $t-1$, and $t-0$). Following the suggestion in (Yu et al. 2017), the temporal convolution operation is conducted without paddings. For that, the length of the output sequences is reduced to a shorter length (i.e., from M to $M - k_t + 1$).

The outputs from the first temporal convolution layer are fed into the spatial graph convolution layer which extracts meaningful patterns and features in the spatial domain. This layer is particularly important when the heat detector from the room of fire origin stops functioning and the determination of flashover occurrence depends on detectors from other compartments. Defferrard et. al (2016) pointed out in their study that the computation of kernel in graph convolution can be expensive, the Chebyshev polynomials approximation is used to facilitate the convolution operation. The outputs of the spatial graph convolution layer (v) are computed as:

$$v(H, A) = ReLU(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H W_g) \tag{2}$$

where A is the adjacency matrix, D is the diagonal degree matrix with $D_{ii} = \sum_j A_{ij}$, and $W_g \in \mathbb{R}^{k_g \times C_i \times C_o^1}$ is the weights in the spatial convolution layer. The kernel size of the graph convolution is k_g. Finally, the complete ST-GonvN is formulated as:

$$f(\mathcal{G}) = f(X, A) = h^2\{v[h^1(X)], A]\} \tag{3}$$

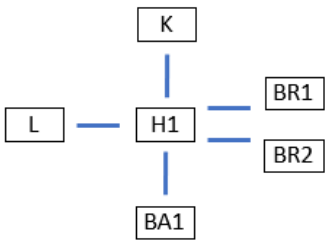
where h^1 and h^2 denote the first and second temporal convolution layers in the block, respectively. Identical kernel size for k_t is used. The propose for the second temporal convolution layer is to obtain higher level temperature information from a larger receptive field which is important for slower fire cases that lead to flashover. Final feature representations are passed into the fully connected output layer to provide the prediction for the potential occurrence of flashover with lead times of 10 s and 30 s. Additional discussion is provided in Sec. 3.5 to obtain the optimal model structure and ablation study (see Sec. 4) is conducted to demonstrate the effectiveness of key components that contribute to the improved outcomes for the model.

3.2 Adjacency Matrix (*A*)

This section provides descriptions for the formulation of adjacency matrices, and there are three separate steps. Firstly, a floor plan of a building structure is converted into a graph representation with nodes and edges. Figure 5a shows the graph representation of the six-compartment attached home (refer to Fig. 1b). As shown in Fig. 5a, each compartment is represented by a node and the corresponding opening or connection between two compartments is represented by an edge. The same procedure is carried out to generate the graph representations for the rest of the sixteen building structures.

In the second step, an adjacency matrix can be obtained for each of the building structures using the graph representations. It can be seen by visual inspection of the seventeen building structures from Appendix C that a building structure can be the combination of fifteen different compartments. Specifically, the combination can be a living room (L), kitchen (K), bedroom1 (BR1), bathroom1 (BA1), hall1 (H1), bathroom2 (BA2), bedroom2 (BR2), bathroom3 (BA3), den (Den), bedroom3 (BR3), family room (Fam), bedroom4 (BR4), hall2 (H2), hall3 (H3), or a dining room (D). Based on this combination, a square matrix with 15 elements can be formulated to describe the compartment relationship within a building structure. Figure 5b presents an adjacency matrix for the six-compartment attached home. The matrix elements, a_{ij} for $i \neq j$, are determined based on the size of opening between two compartments (i.e., the height and the width of a door from a bedroom to the hallway). Since it is expected there will be an exchange of air from one compartment to another compartment in case of fire, the adjacency matrix is symmetric, indicating that temperature information is equally important. In principle, this adjacency matrix is denoted as an undirected graph (Zhou et al. 2020). For the diagonal elements, since the temperature information is crucial to correlate the potential occurrence of flashover, they are taken to be the maximum value of the non-diagonal elements. For the six-compartment building structure, the maximum value is 2.88 m^2. It should be noted that the formulation of the adjacency matrix is based on pre-defined information. However, additional research is on-going to establish a latent learning layer to allow the adjacency matrix to be determined automatically based on the temperature information. In principle, this will help to eliminate the assumptions being made during the formulation of the adjacency matrix and to provide a more robust end-to-end prediction model.

a)



b)

	L	K	BR1	BA1	H1	BA2	BR2	BA3	...		D
L	2.88	0	0	0	2.88	0	0	0			0
K	0	2.88	0	0	1.845	0	0	0			0
BR1	0	0	2.88	0	1.845	0	0	0			0
BA1	0	0	0	2.88	1.845	0	0	0	0		0
H1	2.88	1.845	1.845	1.845	2.88		1.845	0			0
BA2	0	0	0	0	0	0	0	0			0
BR2	0	0	0	0	1.845	0	2.88	0			0
BA3	0	0	0	0	0	0	0	0			0
⋮	0								0		0
											0
D	0	0	0	0	0	0	0	0	0	0	0

Figures 5. A schematic of a) a graph representation and b) an adjacency matric for the six-compartment home (see Fig.1b).

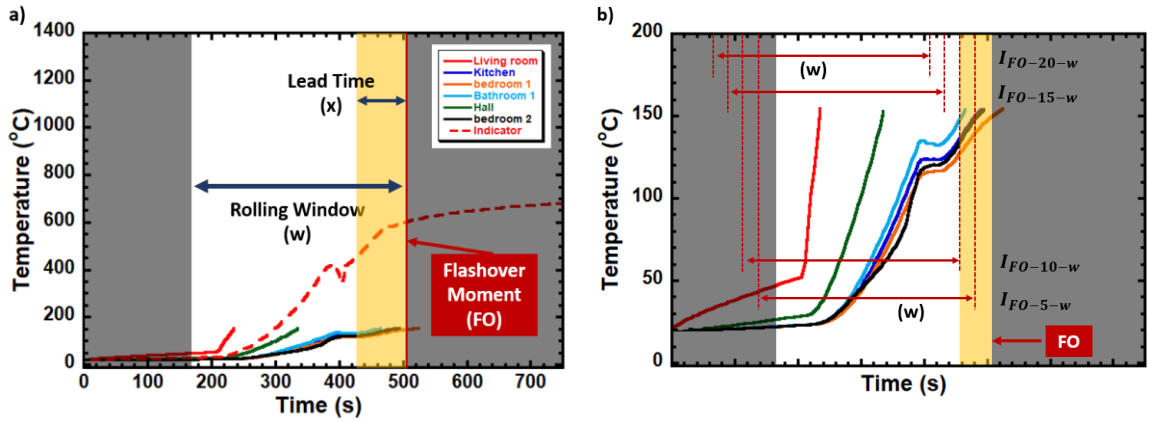
After the adjacency matrices (AM) for the seventeen building structures are formulated, the last step involves the development of a geometric average adjacency matrix. The need of the geometric average adjacency matrix (GAAM) is due to the fact that the GNN algorithm cannot use multiple adjacency matrices for learning. To the best of the authors' knowledge, the current study is the first attempt to overcome the difficulty of having multiple adjacency matrices. Figure 6 shows the GAAM for the selected seventeen structures and each of the matrix elements is determined based on the statistical mean of the seventeen AM. The GAAM is used together with the temperature instances (inputs) to train the generic prediction model.

	L	K	BR1	BA1	H1	BA2	BR2	BA3	Den	BR3	Fam	BR4	H2	H3	D
L	1.00	0.58	0.53	0.52	0.92	0.70	0.65	0	0.64	0	0	0	1.00	0	0.64
K	0.58	1.00	0	0	0.65	0	0	0	0	0	0.85	0	0	0	0.56
BR1	0.53	0	1.00	0.56	0.56	0	0	0	0	0	0	0	0.59	0	0
BA1	0.52	0	0.56	1.00	0.63	0	0	0	0	0	0	0	0.36	0	0
H1	0.92	0.65	0.56	0.63	1.00	0	0.64	0.43	0.00	0.64	0.43	0	0.67	0	0
BA2	0.70	0	0	0	0.00	1.00	0	0	0	0	0	0	0.64	0.43	0
BR2	0.65	0	0	0	0.64	0	1.00	0	0	0	0	0	0.43	0.43	0
BA3	0	0	0	0	0.43	0	0	1.00	0	0	0	0	0	0	0
Den	0.64	0	0	0	0.00	0	0	0	1.00	0	0	0	0.53	0	0
BR3	0	0	0	0	0.64	0	0	0	0	1.00	0	0	0.85	0	0
Fam	0	0.85	0	0	0.43	0	0	0	0	0	1.00	0	0	0	0
BR4	0	0	0	0	0	0	0	0	0	0	0	1.00	0.43	0	0
H2	1.00	0	0.59	0.36	0.67	0.64	0.43	0	0.53	0.85	0	0.43	1.00	1.00	0.43
H3	0	0	0	0	0	0.43	0.43	0	0	0	0	0	1.00	1.00	0
D	0.64	0.56	0	0	0	0	0	0	0	0	0	0	0.43	0	1.00

Figure 6. The geometric average adjacency matrix for the seventeen building structures.

3.3 Node Attributes (*X*)

In this study, the data consists of temperature readings from the heat detectors, and the objective of the model is to predict if there will be a potential occurrence of flashover from a particular compartment within a building structure in the future. Considering a building structure with *N* heat detectors, the temperature data is denoted as $T = [T_0, T_5, \dots, T_{FO}]$. The data sampling interval is 5 s and the notation, *FO*, is the time stamp for the transition to flashover. In order to allow firefighters to have sufficient time to leave the compartment or take action, the model should provide the predictions of a potential occurrence of flashover ahead of time. For that, lead time, *x*, is considered. Since it will take 10 s for firefighters to travel about 3 m due to movement limitations and poor visibility (Dunn 2015), two lead times are selected, and they are 10 s and 30 s. Figure 7a provides an illustration of the lead time being applied to the medium growth fire case (refer to Figs. 3).



Figures 7. An illustration of a) using a rolling window with a lead time for a fire case and b) non-flashover and flashover instances using a window size of w. Note that the lead time region is not in scale.

As shown in Fig. 7a, a sliding window is applied to construct instances from raw temperature data. An instance from a fire event[1] is denoted as $I_i = \{S_i^1, S_i^2, \dots, S_i^N\}$ where $S_i^j = (T_t^j, \dots, T_{t+w}^j)$ with i as the first time step of the sliding window and w to be the window size. Motivated by Menon et. al (2017) and Arunabha Roy (2022), in order to examine the effect of window size on model performance, six window sizes are chosen as 60s, 90 s, 120 s, 180 s, 240 s, and 300 s. The maximum window size is selected due to limits associated with the existing fire protection devices. To account for heat detector failure at 150 °C, a failure moment (t_b^j) for signal S^j is identified. If time $t \geq t_b^j$, the temperature reading, T_t^j, is replaced by a value of 0 °C, representing a loss of temperature signal. A masking layer is applied during the training process to neglect the zero values. Extracting all the I_i from all fire events, the instance set of $\{I_0^1, I_1^1, \dots, I_i^1, \dots, I_0^K, I_1^K, \dots, I_i^K\}$ is obtained to be the node attributes (X) where K is the number of fire cases, and the exact number of instances being used to train the current model will be provided in the following subsection. It should be noted that padding is used to make the node attributes (i.e., a vector with 15 elements) consistent with the adjacency matrix.

3.4 Labelling

Each instance is labelled to form the data samples[2]. Since the model is to predict the potential occurrence of flashover, the instance is either labelled as *flashover* or *non-flashover* based on the upper gas layer temperature. In this study, the threshold for the onset of flashover is selected to be 600 °C (Peacock et al. 2019; Reneke 2013). Figure 7b provides an illustration of flashover and non-flashover instances. For predictions with a lead time of 10 s, there are two flashover instances before the flashover condition is met and they are $I_{FO-10-w}$ and I_{FO-5-w}. Given the fact that only about three-fifth of the fire cases (i.e., 41,129 out of 78,000 cases) have flashover and a majority of the instances from a flashover fire case are non-flashover, the current dataset has a data imbalance between flashover and non-flashover instances. In order to maintain data balance, two non-flashover instances are selected, and they are taken to be $I_{FO-20-w}$ and $I_{FO-15-w}$. In total, there are 164,516 (i.e., 2 x 2 x 41,129) instances from 41,129 fire cases that have met the flashover condition. For predictions with a lead time of 30 s, since there are six non-flashover and flashover instances (12 for a fire), it yields a total of 493,548 instances.

3.5 Training and Testing

The dataset is divided into three subsets: training (60 %), validation (20 %), and testing (20 %) subsets. Specifically, a set of 8114 fire events worth of data samples are assigned to both validation and testing sets. The remaining 24681 (41129 - 16448) fire events are used for the training set. In order to facilitate the training process, the dataset is divided into batches, and the batch size is selected to be 50. Using the RMSProp with an initial learning rate of 10^{-3}, the neural network weights and biases from the above equations are updated accordingly. Following the algorithms in Fu et. al (2021) and Xiao et. al (2021b), softmax cross-entropy is used as the loss function to compute the difference between the ground-truths (y_j) and the corresponding prediction results (p_j) and the loss function for binary classifications is written as:.

$$Loss = -\frac{1}{N}\sum_{j=1}^{N} -\left[y_j \cdot log(p_j) + (1 - y_j) \cdot log(1 - p_j)\right] \quad (4)$$

where N is the number of instances.

Early-stopping with a patience (Abadi et al. 2016) of 5 is used, and Fig. 8 shows the validation loss and accuracy for a training process. As seen in the figure, even though the accuracy still increases slightly after epoch 43 and since the loss has not decreased for 5 consecutive epochs, the training stops to avoid overfitting. Numerical studies are conducted to select the optimized ST-GonvN model structure, and it is found that the model with a node size of 1, 32, and 16 for the temporal, spatial, and temporal convolutional

[1] A fire event is considered as a fire case, and there are 78000 fire cases for seventeen different building structures in this study.

[2] Based on Hamins et. al (2015), firefighters need actionable information. This is because the fire situation changes rapidly and firefighters typically do not have enough time to interpret the model outputs. For that, the current study focuses on a classification task.

layer, respectively, provides the desired model performance. The kernel size of 3 is used for the spatial and temporal convolutions. With this model configuration, the ratio of training data to trainable parameters is about 27.7 times. The number of parameters is provided in below section. Appendix D presents a comparison of the results for 35 different ST-GonvN model structures.

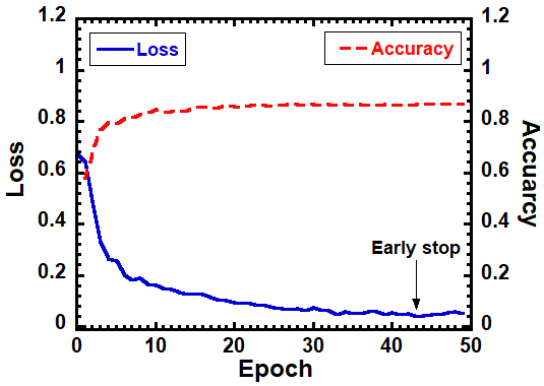


Figure 8. Validation loss and accuracy for a model training and the illustration of using the early-stopping with a patience of 5 to avoid overfitting.

4 RESULTS

4.1 Model Performance of FlashNet

Table 4 shows the results for the prediction of potential flashover occurrence with a lead time of 10 s from a rolling window size of 300 s. FlashNet is benchmarked against five state-of-te-art flashover prediction models. The baseline models include i) SVM – a support vector machine using a linear kernel (Wang et. al 2020), ii) MLP – a feedforward multiple-layer perceptron (Yuen et. al 2006), iii) LSTM – a two-layer long short-term memory (Zhang et. al 2022), iv) BiLSTM-ATT – a bi-directional LSTM with attention mechanism (Fu et. al 2021), and v) CNN – a three-layer standard convolutional neural network. Each model is fine-tuned to obtain optimal model performance without overfitting. The following metrics: accuracy, precision, recall, and F1 score are used to evaluate the model performance. The mathematical expressions are given as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{5a}$$

$$Precision = \frac{TP}{TP + FP} \tag{5b}$$

$$Recall = \frac{TP}{TP + FN} \tag{5c}$$

$$F1 = 2 * \frac{Recall * Precision}{Recall + Precision} \tag{5d}$$

where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively. Ture positive is defined as correction prediction of flashover conditions when the label is flashover and false positive is defined as misprediction of flashover conditions when the label is flashover. Basically, these metrics offer additional insights about incorrect predictions, either positive or negative, and the effects of data imbalance.

As shown in Table 4, FlashNet outperforms the existing machine learning based flashover prediction models. In general, it achieves an overall accuracy of about 85.2 %. Also, the scores from precision, recall, and F1 suggest that FlashNet is a well-balanced model which minimizes the false positive and the false negative. The main reason why FlashNet improves the model performance is that the nature for temperature data in flashover predictions accounting for different building structures with a wide range of fire and vent opening conditions is better suited for the modelling assumptions about temporal and spatial dependencies.

In terms of training time, most of models, except ANN and the linear SVM, require about 30 minutes for convergence. For testing, models generally take about 3 seconds to make predictions for testing subset. On average, one prediction uses about 3e-5 seconds-per-instance (i.e., 3/(493,548*0.2)).

Table 4. Baseline comparison against five existing machine learning based flashover prediction models.

Methods	**Accuracy**	**Precision**	**Recall**	**F1**	**Training (s)**	**Testing (s)**	**Num. of Params**
SVM	53.6%	52.8%	66.6%	58.9%	37.5	0.01	NA
MPL	62.1%	49.4%	74.3%	59.3%	710.6	2.42	9,457
LSTM	66.6%	75.7%	49.7%	60.1%	1594.1	2.87	11,602
BiLSTM-ATT	70.0%	66.6%	81.2%	73.2%	1834.8	2.93	12,217
CNN	79.8%	85.0%	74.8%	79.5%	1351.0	2.38	13,322
FlashNet	85.2%	83.5%	87.8%	85.6%	1657.3	2.91	10,687

Ablation study is conducted to examine the effectiveness of key components that contribute to the improved outcomes for FlashNet. The full model of FlashNet is compared with four model variations:

i) w/o GCNN – FlashNet without the graph convolution layer and it is being replaced with a linear layer,
ii) ii) w/o TCNN – FlashNet without both of the temporal convolution layers and they are replaced with linear layers,
iii) iii) w/o 1st TCNN – replacing the first temporal convolution layer with a linear layer, and
iv) iv) w/o 2nd TCNN – replacing the second temporal convolution layer with a linear layer.

Table 5 shows the accuracy, precision, recall and F1 scores for each of the models. The introduction of graph convolution layer significantly improves the results as it allows the model to learn the limited temperature information from different compartments. The effect of the sandwich structure of the two TCNN layers is evident as well; it validates that the use of two TCNN facilitates the selection of useful information from the time domain.

Table 5. Ablation study for FlashNet.

Methods	FlashNet	w/o GCNN	w/o TCNN	w/o 1st TCNN	w/o 2nd TCNN
Accuracy	**85.2%**	80.3%	83.7%	84.4%	84.2%
Precision	**83.5%**	81.8%	81.6%	82.6%	83.5%
Recall	**87.8%**	77.9%	86.8%	87.3%	85.3%
F1	**85.6%**	79.8%	84.1%	84.9%	84.4%

Table 6 shows the results of FlashNet for the prediction of potential flashover occurrence with two different lead times (10 s and 30 s) and six rolling window sizes (300 s, 240 s, 180 s, 120 s, 90 s, and 60 s). As shown in Table 6, the best overall model accuracy is ~ 92.1 % using temperature information from a 300 s time window with a prediction lead time of 30 s. For this case, the precision and recall scores are ~ 91.2 % and ~ 92.4 %, respectively. Comparing precision and recall scores, it can be seen that there are more false positives than false negatives. And in practical firefighting applications, it is desired to minimize the number of false negatives. However, given the fact that only limited temperature information (i.e., up to 150 °C and no prior knowledge about the exact building structure) is used as inputs for predictions, the current results are very encouraging. In Table 6, it can also be seen that model performance tends to decrease with a smaller time window and/or a smaller lead time. Physically, this observation is expected because less temperature information is used with a smaller time window and a smaller lead time yields a ~ 66% reduction for training and testing instances. It is interesting to note that there is about 3 % drop in model accuracy for a large reduction of the time window size from 300 s to 60 s for lead time of 30 s. This suggests that a long duration of temperature information might not be needed.

Table 6. Overall FlashNet model performance for 17 different homes.

Time window	Lead time	Accuracy	Precision	Recall	F1
300 s	**10 s**	85.2%	83.5%	87.8%	85.6%
	30 s	92.1%	91.2%	92.4%	91.8%
240 s	**10 s**	84.4%	82.1%	86.5%	84.2%
	30 s	91.8%	90.7%	92.9%	91.8%
180 s	**10 s**	83.1%	81.5%	85.6%	83.5%
	30 s	91.9%	90.6%	91.3%	90.9%
120 s	**10 s**	81.1%	76.4%	84.1%	80.0%
	30 s	91.5%	85.8%	91.7%	88.7%
90 s	**10 s**	80.4%	82.6%	78.6%	80.6%
	30 s	89.4%	85.8%	93.8%	89.6%
60 s	**10 s**	79.8%	80.7%	77.2%	78.9%
	30 s	89.1%	85.8%	93.8%	89.6%

In order to understand the model performance as a function of instances for different homes, additional results are obtained. Table 7 shows the accuracy of the non-flashover instances (20 s before flashover and 15 s before flashover) and the flashover instances (10 s before flashover and 5 s before flashover) with a prediction lead time of 10 s for 17 different homes using a 300 s time window. An important observation is found, and that is the prediction accuracy for instances associated with 20 s before flashover and 5 s before flashover (5 s BFO) are much higher than the prediction accuracy for instances associated with 15 s BFO and 10 BFO. It can be seen that the average accuracy for 15 s BFO and 10 BFO is about 11 % lower. In principle, this result tendency is due to the fact that the temperature signals used to make predictions from these two instances are relatively similar but the corresponding labels are completely different (i.e., 15 s BFO as non-flashover and 10 s BFO as flashover. Similar behaviors are also observed in Table 8 which depicts the accuracy of the non-flashover and the flashover instances with a prediction lead time of 30 s for 17 different homes using a 300 s time window. The average accuracy for 35 s BFO (non-flashover) and 30 BFO (flashover) is significantly lower. However, as seen in both tables, the model predictions become more reliable for other instances, and the average accuracy is about 94 % and 98 % for predictions with a lead time of 10 s and 30 s, respectively. To provide more reliable actionable information to the firefighters about the potential occurrence of flashover, a larger lead time is preferred.

Table 7: Accuracy for non-flashover and flashover instances with a lead time of 10 s among 17 different homes using a 300 s time window.

Accuracy for Different Homes (%)																		
Instances		**# 1**	**# 2**	**# 3**	**# 4**	**# 5**	**# 6**	**# 7**	**# 8**	**# 9**	**# 10**	**# 11**	**# 12**	**# 13**	**# 14**	**# 15**	**# 16**	**# 17**	***AVG.***
20 s	**before flashover**	80.0	100.0	93.6	93.5	94.2	94.1	93.4	94.8	93.8	95.6	94.5	94.3	94.3	94.0	94.7	93.9	94.3	*93.7*
15 s		80.0	89.9	77.9	80.8	79.8	81.3	80.7	81.7	80.7	81.2	80.1	80.6	79.7	80.7	82.2	81.8	82.4	*81.3*
10 s		100.0	80.7	82.6	79.1	82.2	83.0	84.7	83.2	81.6	82.8	81.2	81.0	82.6	80.6	80.9	78.7	82.7	*82.8*
5 s		100.0	90.2	94.5	92.7	94.1	91.9	92.3	92.3	93.1	92.3	93.4	92.1	95.9	93.1	93.6	92.1	93.1	*93.3*

Table 8: Accuracy for non-flashover and flashover instances with a lead time of 30 s among 17 different homes using a 300 s time window.

Accuracy for Different Homes (%)																			
Instances		**# 1**	**# 2**	**# 3**	**# 4**	**# 5**	**# 6**	**# 7**	**# 8**	**# 9**	**# 10**	**# 11**	**# 12**	**# 13**	**# 14**	**# 15**	**# 16**	**# 17**	***AVG.***
60 s	**before flashover**	95.6	99.3	95.1	96.0	96.0	95.9	96.5	96.8	96.8	97.1	96.8	96.7	96.4	96.3	96.9	96.6	96.7	*96.6*
55 s		95.6	99.2	95.0	95.2	95.6	94.6	96.4	96.2	96.1	96.7	95.6	95.4	96.5	95.9	95.9	96.3	96.1	*96.0*
50 s		94.3	99.0	93.9	94.6	95.0	93.3	95.2	94.7	95.5	96.3	94.9	94.4	96.2	94.5	95.9	95.8	95.5	*95.2*
45 s		94.1	98.9	93.1	94.6	95.0	94.2	94.9	95.2	95.8	95.8	94.5	94.8	95.2	94.3	95.4	95.6	95.2	*95.1*
40 s		91.9	99.5	90.8	94.0	93.8	94.1	95.1	95.0	94.3	96.1	94.5	94.6	95.2	93.1	95.7	94.0	94.8	*94.5*
35 s		72.2	82.8	71.1	84.1	76.2	81.0	84.8	85.0	75.7	85.5	83.9	85.1	85.3	84.3	85.1	75.6	84.6	*81.3*
30 s		81.7	78.9	81.0	78.0	83.0	77.1	79.7	76.7	83.2	79.9	79.1	78.1	80.0	79.1	76.8	83.2	79.9	*79.7*
25 s		97.3	95.1	96.0	95.2	98.2	95.8	96.5	95.5	97.9	96.5	95.9	95.3	96.6	95.8	95.9	98.4	97.1	*96.4*
20 s		98.7	99.2	97.3	96.9	98.9	97.7	96.7	96.8	98.0	97.9	96.7	96.9	97.9	97.1	96.3	99.0	97.5	*97.6*
15 s		98.6	99.5	98.8	97.3	99.1	97.9	97.4	97.5	98.1	97.8	97.5	97.8	98.0	97.1	96.5	99.2	97.5	*98.0*
10 s		99.1	99.6	98.6	97.7	99.2	97.9	97.7	97.8	98.3	98.0	97.9	97.9	98.0	97.3	96.9	99.2	97.6	*98.2*
5 s		99.1	99.6	98.4	98.4	99.4	98.0	97.7	98.4	98.6	98.2	98.1	97.9	98.2	97.9	97.7	99.3	97.7	*98.4*

4.2 The Effect of Geometric Average Adjacency Matrix

The effect of adjacency matrix to the model performance has never been explored because the use of GNNs for studies related to flashover and/or other hazard predictions is limited in the fire research and other engineering communities. In order to understand the effectiveness of the geometric average adjacency matrix, a case study is conducted. Since a baseline adjacency matrix is not available in existing literature, an all-one adjacency matrix is used to carry out this study. The all-one adjacency matrix has a dimension of 15 x 15 and each of the matrix elements is equal 1. Physically, the all-one adjacency matrix includes the inherent assumption that the effect of the actual opening areas between different compartments is negligible.

Two models are obtained separately using the geometric average adjacency matrix (GAAM) and the all-one adjacency matrix (All-One AM). Data with a lead time of 10 s and a time window of 90 s are utilized to facilitate the model training process. The data is divided into three subsets (i.e., a training, validation, and a testing set) with the data ratio of 0.6, 0.2, and 0.2, respectively. Both models are trained using a smaller neural network structure with node size of 1, 16, and 32 for the three different convolutional layers.

Figure 9 shows the validation loss and accuracy for models using All-One AM and GAAM with the node size of 1-16-32. As shown in the figure, the use of GAAM yields a notably lower loss (or errors), and the relative difference at epoch 25 is about 29 %. In terms of accuracy, the model using GAAM has an overall accuracy of about 82.8 %. In comparison to the model using All-One AM, the model performance is about 7.7 % more accurate. Additional results are obtained to investigate the model performance in the testing set. Tables 9 show the prediction accuracy of non-flashover and flashover instances for the two different models. As seen in Table 9a, the model using of All-One AM tends to have lower prediction accuracy in flashover instances. For 10 s BFO instances, the accuracy is only about 56 %. For 5 s BFO instances, although the accuracy reaches about 84 %, the model performance is still significantly lower than the model using GAAM, and the difference is more than 10 %. Another important observation seen in Fig. 9 is that model complexity will need to be largely enhanced in order to have better model performance. Using the All-One AM, the validation loss is reduced, and accuracy becomes higher when the node size increases to 1-32-64. Even though better model performance is achieved, the increase in model complexity is not necessary. The drawback is that the required training time is increased by ~ 19 % and the numerical efficiency (i.e., making a prediction) is reduced by ~ 32.2 %. Therefore, a properly designed adjacency matrix, such as GAAM, should be used to develop a robust model.

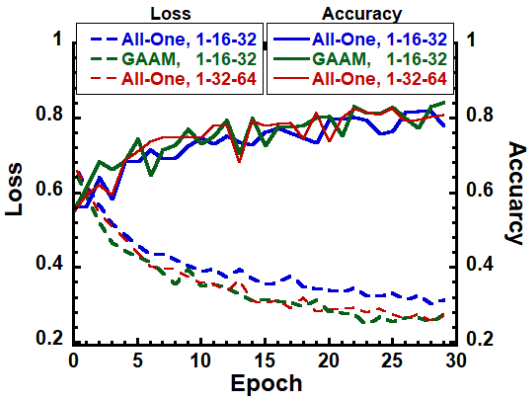


Figure 9. Validation loss and accuracy for three different STGCN models using: i) all-one AM with 1-16-32 nodes, ii) GAAM with 1-16-32 nodes, and iii) all-one AM with 1-32-64 nodes.

Table 9a: Accuracy for the model using All-One AM.

Home	Accuracy			
	20 s	15 s	10 s	5 s
	before flashover			
#1	81.1%	79.0%	51.4%	67.8%
#2	91.7%	87.5%	45.8%	70.8%
#3	91.1%	86.7%	59.0%	87.7%
#4	93.2%	86.3%	57.6%	85.1%
#5	91.8%	81.8%	55.6%	82.2%
#6	94.3%	86.8%	52.3%	79.9%
#7	95.5%	87.5%	59.3%	88.5%
#8	95.0%	87.5%	57.4%	85.0%
#9	93.7%	87.4%	54.6%	84.3%
#10	95.9%	88.0%	59.4%	89.7%
#11	94.2%	87.5%	58.4%	87.2%
#12	93.8%	87.0%	57.5%	86.3%
#13	95.7%	88.0%	59.6%	89.6%
#14	94.4%	87.7%	58.0%	87.0%
#15	95.1%	86.6%	55.3%	82.7%
#16	92.9%	85.6%	51.8%	84.1%
#17	95.2%	87.3%	59.2%	88.5%
AVG.	*93.2%*	*86.4%*	*56.0%*	*83.9%*

Table 9b: Accuracy for the model using GAAM.

Home	Accuracy			
	20 s	15 s	10 s	5 s
	before flashover			
#1	84.9%	78.8%	81.1%	86.6%
#2	91.7%	83.3%	79.2%	91.0%
#3	92.4%	80.4%	78.8%	92.7%
#4	91.4%	80.9%	77.7%	95.2%
#5	93.1%	78.8%	77.4%	94.2%
#6	92.0%	78.2%	79.3%	92.0%
#7	94.3%	83.5%	78.5%	96.6%
#8	92.8%	79.9%	78.0%	94.4%
#9	93.0%	82.1%	78.0%	94.8%
#10	95.0%	84.4%	79.2%	96.9%
#11	93.0%	82.2%	77.3%	96.1%
#12	92.4%	81.4%	76.7%	95.8%
#13	94.7%	84.4%	79.5%	96.7%
#14	93.1%	82.4%	77.1%	95.9%
#15	93.0%	79.9%	79.9%	93.7%
#16	93.7%	81.0%	79.4%	95.3%
#17	94.1%	83.5%	78.6%	96.5%
AVG.	*92.6%*	*81.5%*	*78.6%*	*94.4%*

5 CONCLUSIONS AND FUTURE WORK

This paper proposes a novel fire scene-agnostic model (FlashNet) for early flashover prediction based on graph neural networks. As a first attempt to tackle building structure invariance in flashover prediction, different channels of the sensor inputs are transformed to a graph according to the building structure and presented to the graph neural networks. A geometric average adjacency matrix is formulated based on multiple common home structures. With this adjacency matrix, FlashNet effectively captures the spatial and temporal dependencies from the multivariate temperature data. To train a well-performing FlashNet, over 41,000 flashover cases with a wide range of fire conditions are used from 17 different home structures with different floorplans. Results show that FlashNet can predict flashover in a realistic setting, where the available temperature information is limited (up to 150 °C) and no prior knowledge is accessible about building structure, fire locations, or vent conditions. The overall accuracy is about 92.1 % with a possible forecast lead time of 30 seconds. With the current model structure, FlashNet can provide rapid predictions and a single prediction takes less than a second. Parametric studies are also conducted to understand the impact of lead time, time window, and the formulation of adjacency matrix on the overall model performance. The evaluation suggests that a decrease in lead time and/or time window generally yields a lower model accuracy and a well formulated adjacency matrix, such as GAAN, provides a more robust model.

Even though FlashNet is numerically efficient and accurate, there can still be improvement. The first improvement can be made to adapt a graph learning layer. This is due to the fact that the current model requires a pre-defined graph and the pre-defined graph is determined based on the authors' domain expertise on the fundamental understanding of fire dynamics and the observations made from full-scale fire experiments. Although results show that the geometric average adjacency matrix provides well-established relationships to describe compartment connections, the inclusion of a graph learning layer is likely to provide better modeling generality for more complex data behavior and/or to establish a more robust adjacency matrix to facilitate model training.

Another possible improvement is that the modeling structure can be expanded to make use of additional sensor signals available from existing smart homes, such as temperature from thermostats and carbon

monoxide and carbon dioxide signals from fire protection devices, to address other fire events, such as occupant tenability. The use of multi-sensor signals from one compartment and the capability of providing multiple outputs/predictions will be beneficial to search and rescue tasks for firefighters. Additional efforts are on-going to achieve the abovementioned improvements. Taken together, this study and the future work will contribute to the development of a real-life flashover prediction model and paves the way to data-driven firefighting with smart systems.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. and Kudlur, M., 2016. {TensorFlow}: A System for {Large-Scale} Machine Learning. In 12th USENIX symposium on operating systems design and implementation (OSDI 16) (pp. 265-283).

Babrauskas, V., 1980. Estimating room flashover potential. Fire Technology, 16(2), pp.94-103.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B. and Anderson, H., 2018. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228.

Campbell, R., Evarts, B. and Molis, J.L., 2019. United States Firefighter Injuries in 2018. NFPA Research, pp.1-27. National Fire Protection Association, USA.

Cleary, T.G. 2014. Improving Smoke Alarm Performance: Justification for New Smoldering and Flaming Test Performance Criteria. National Institute of Standards and Technology. TN 1837. Gaithersburg, MD.

Dexters, A., Leisted, R.R., Van Coile, R., Welch, S., Jomaas, G., 2021, Testing for knowledge: Application of machine learning techniques for prediction of flashover in a 1/5 scale ISO 13784-1 enclosure, Fire and Materials. 45, 708–719. https://doi.org/10.1002/fam.2876.

Dunn, V. 2015. Fire Engineering Books: Safety and Survival on the Fireground. Pennwell Books. United States.

Fahy, R.F., Petrillo, J.T. and Molis, J.L., 2020. Firefighter Fatalities in the US-2019. NFPA Research, pp.1-26. National Fire Protection Association, USA.

Fu, E.Y., Tam, W.C., Wang, J., Peacock, R., Reneke, P., Ngai, G., Leong, H.V. and Cleary, T., 2021, May. Predicting Flashover Occurrence using Surrogate Temperature Data. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 17, pp. 14785-14794).

Garrity, D.J. and Yusuf, S.A., 2021. A predictive decision-aid device to warn firefighters of catastrophic temperature increases using an AI-based time-series algorithm. Safety science, 138, p.105237.

Hamins, A.P., Bryner, N.P., Jones, A.W. and Koepke, G.H., 2015. Research roadmap for smart fire fighting. NIST SP 1191. National Institute of Standards and Technology, Gaitersburg, MD.

Hodges, J.L., Lattimer, B.Y. and Luxbacher, K.D., 2019. Compartment fire predictions using transpose convolutional neural networks. Fire Safety Journal, 108, p.102854.

Hurley, M.J., Gottuk, D.T., Hall Jr, J.R., Harada, K., Kuligowski, E.D., Puchovsky, M., Watts Jr, J.M. and Wieczorek, C.J. eds., 2015. SFPE handbook of fire protection engineering. Quincy, Massachusetts.

Jahn, W., 2017. Using suppression and detection devices to steer CFD fire forecast simulations. Fire Safety Journal, 91, pp.284-290.

Kim, H.J. and Lilley, D.G., 2002. Flashover: A study of parameter effects on time to reach flashover conditions. Journal of propulsion and power, 18(3), pp.669-673.

Koo, S.H., Fraser-Mitchell, J. and Welch, S., 2010. Sensor-steered fire simulation. Fire Safety Journal, 45(3), pp.193-205.

Menon, R., Di Caterina, G., Lakany, H., Petropoulakis, L., Conway, B.A. and Soraghan, J.J., 2017. Study on interaction between temporal and spatial information in classification of EMG signals for myoelectric prostheses. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25(10), pp.1832-1842.

McCaffrey, B.J., Quintiere, J.G. and Harkleroad, M.F., 1981. Estimating room temperatures and the likelihood of flashover using fire test data correlations. Fire Technology, 17(2), pp.98-119.

McKinnon, M., Weinschenk, C. and Madrzykowski, D. 2020. Modeling Gas Burner Fires in Ranch and Colonial Style Structures. Underwriters Laboratories Firefighter Safety Research Institute. Columbia, MD.

Mitler, H.E. and Steckler, K.D., 1995. Wall fires and the approach to flashover in an enclosure. Safety science, 20(1), pp.71-78.

Nasiri, S., Khosravani, M.R. and Weinberg, K., 2017. Fracture mechanics and mechanical fault detection by artificial intelligence methods: A review. Engineering Failure Analysis, 81, pp.270-293.

NFPA 72: National Fire Alarm and Signaling Code, 2002 Edition. In NFPA National Fire Codes Online. Retrieved from http://codesonline.nfpa.org

Overholt, K. J., & Ezekoye, O. A., 2012. Characterizing heat release rates using an inverse fire modeling technique. Fire Technology, 48(4), 893-909.

Peacock, R.D., Forney, G.P. and Reneke, P.A., 2015. CFAST—consolidated fire and smoke transport (version 7)—volume 3: verification and validation guide. NIST Technical Note 1889v3 (National Institute of Standards and Technology, Gaithersburg, MD, 2015).

Peacock, R.D., Reneke, P.A., Bukowski, R.W. and Babrauskas, V., 1999. Defining flashover for fire hazard calculations. Fire Safety Journal, 32(4), pp.331-345

Persily, A., Musser, A. and Leber, D.D., 2006. A collection of homes to represent the US housing stock. US NISTIR 7330. National Institute of Standards and Technology, Gaithersburg, MD.

Pomeroy, A. T. (2010). Analysis of the effects of temperature and velocity on the response time index of heat detectors. University of Maryland. (MS Dissertation).

Price, M.D., 2014. Using inverse fire modeling with multiple input signals to obtain heat release rates in compartment fire scenarios (Doctoral dissertation, University of Maryland, College Park).

Rao, Q. and Frtunikj, J., 2018, May. Deep learning for self-driving cars: Chances and challenges. In Proceedings of the 1st International Workshop on Software Engineering for AI in Autonomous Systems (pp. 35-38).

Reneke, P.A., 2013. Towards Smart Fire Panels. NIST TN 1780. National Institute of Standards and Technology, Gaitersburg, MD.

Reneke, P.A., Bruns, M., Gilbert, S.W., MacLaren, C.P., Peacock, R.D., Cleary, T.G. and Butry, D.T. 2019. Towards a Process to Quantify the Hazard of Fire Protection Design Alternatives. NIST TN-2041. National Institute of Standards and Technology, Gaithersburg, MD.

Reneke, P.A., Peacock, R.D., Gilbert, S.W. and Cleary, T.G., 2021. CFAST–Consolidated Fire and Smoke Transport (Version 7) Volume 5: CFAST Fire Data Generator (CData). NIST TN 1889v5. National Institute of Standards and Technology, Gaitersburg, MD.

Roy, A.M., 2022. An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces. Biomedical Signal Processing and Control, 74, p.103496.

Swazinna, P., Udluft, S. and Runkler, T., 2021. Overcoming model bias for robust offline deep reinforcement learning. Engineering Applications of Artificial Intelligence, 104, p.104366.

Sugano, Y., Matsushita, Y. and Sato, Y., 2014. Learning-by-synthesis for appearance-based 3d gaze estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1821-1828).

Tam, W.C., Fu, E.Y., Peacock, R., Reneke, P., Wang, J., Li, J. and Cleary, T., 2020. Generating synthetic sensor data to facilitate machine learning paradigm for prediction of building fire hazard. Fire Technology, pp.1-22.

Wang, J., Tam, W.C., Jia, Y., Peacock, R., Reneke, P., Fu, E.Y. and Cleary, T., 2021. P-Flash–A machine learning-based model for flashover prediction using recovered temperature data. Fire Safety Journal, 122, p.103341.

Xiao, Z., Xu, X., Xing, H., Song, F., Wang, X. and Zhao, B., 2021a. A federated learning system with enhanced feature extraction for human activity recognition. Knowledge-Based Systems, 229, p.107338.

Xiao, Z., Xu, X., Xing, H., Luo, S., Dai, P. and Zhan, D., 2021b. RTFN: a robust temporal feature network for time series classification. Information Sciences, 571, pp.65-86.

Wu, X., Park, Y., Li, A., Huang, X., Xiao, F. and Usmani, A., 2021. Smart detection of fire source in tunnel based on the numerical database and artificial intelligence. Fire Technology, 57(2), pp.657-682.

Yu, B., Yin, H. and Zhu, Z., 2017. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875.

Zhang, T., Wang, Z., Wong, H.Y., Tam, W.C., Huang, X., Xiao, F. (2022) Real-time Forecast of Compartment Fire and Flashover based on Deep Learning. Fire Safety Journal, 130, p.103579.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C. and Sun, M., 2020. Graph neural networks: A review of methods and applications. AI Open, 1, pp.57-81.

Appendix A: Model Validation Against Full-Scale Experiments

In order to make sure that CData can be used to generate the corresponding temperature data for different fire scenarios, model validation is carried out. Specifically, temperature measurements obtained from two full-scale gas burner experiments reported in (McKinnon et al. 2020), where a fire initiated in the living room within the single-story residential structure is used to benchmark the synthetic data. The fire location and the heat release rate (HRR) of the gas burner fires are the same for the two tests (Exp 1 and Exp 2). Yet, opening conditions of each of doors and windows are different. It should be noted that natural gas burners are used in these pre-run experiments. The reason is that the HRR of the fire can be fully controlled by regulating how much natural gas is being burned. By doing so, the simulation conditions and the experimental conditions can be adjusted to have identical settings.

Figures A1 show the upper gas layer temperature profiles for the two experiments. The blue solid lines represent the synthetic temperature data obtained from CFAST. The red dash lines are the estimated upper gas layer temperature for the experiments. It can be seen that the magnitudes and trends of the temperature profiles match the experimental data for different vent opening events. This observation indicates that CFAST, the simulation engine of CData, is capable of capturing both the corresponding effect of fire and vent openings in the single-story multi-compartment structure. In terms of uncertainty, the absolute root mean squared error is about 30 °C and 10 °C for Exp 1 and Exp 2, respectively. Comparisons are also made for the other compartments and the overall agreement is very good. Therefore, it can be expected that the generated temperature data can be reliably used for model development.

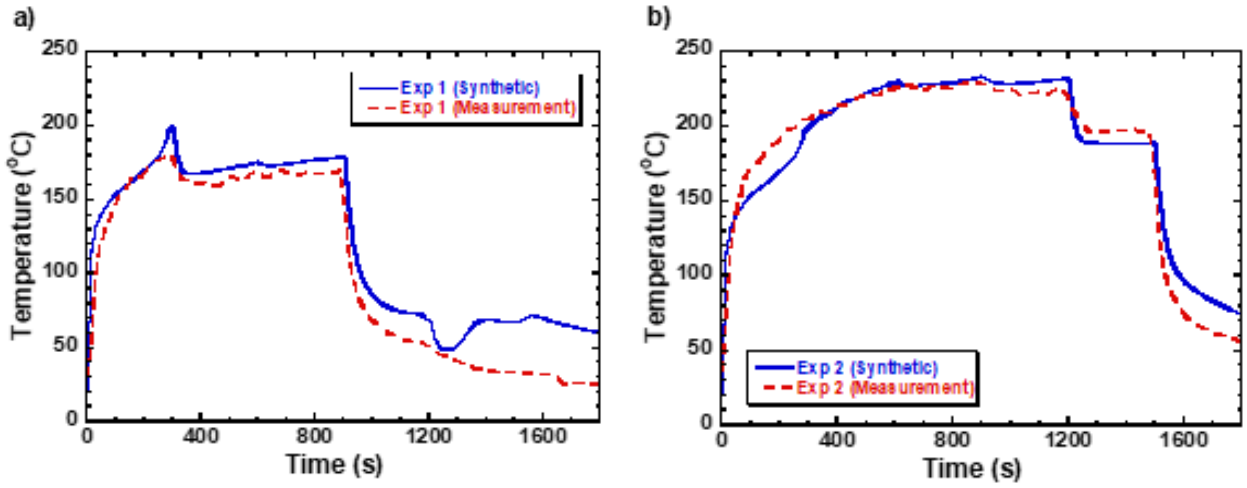
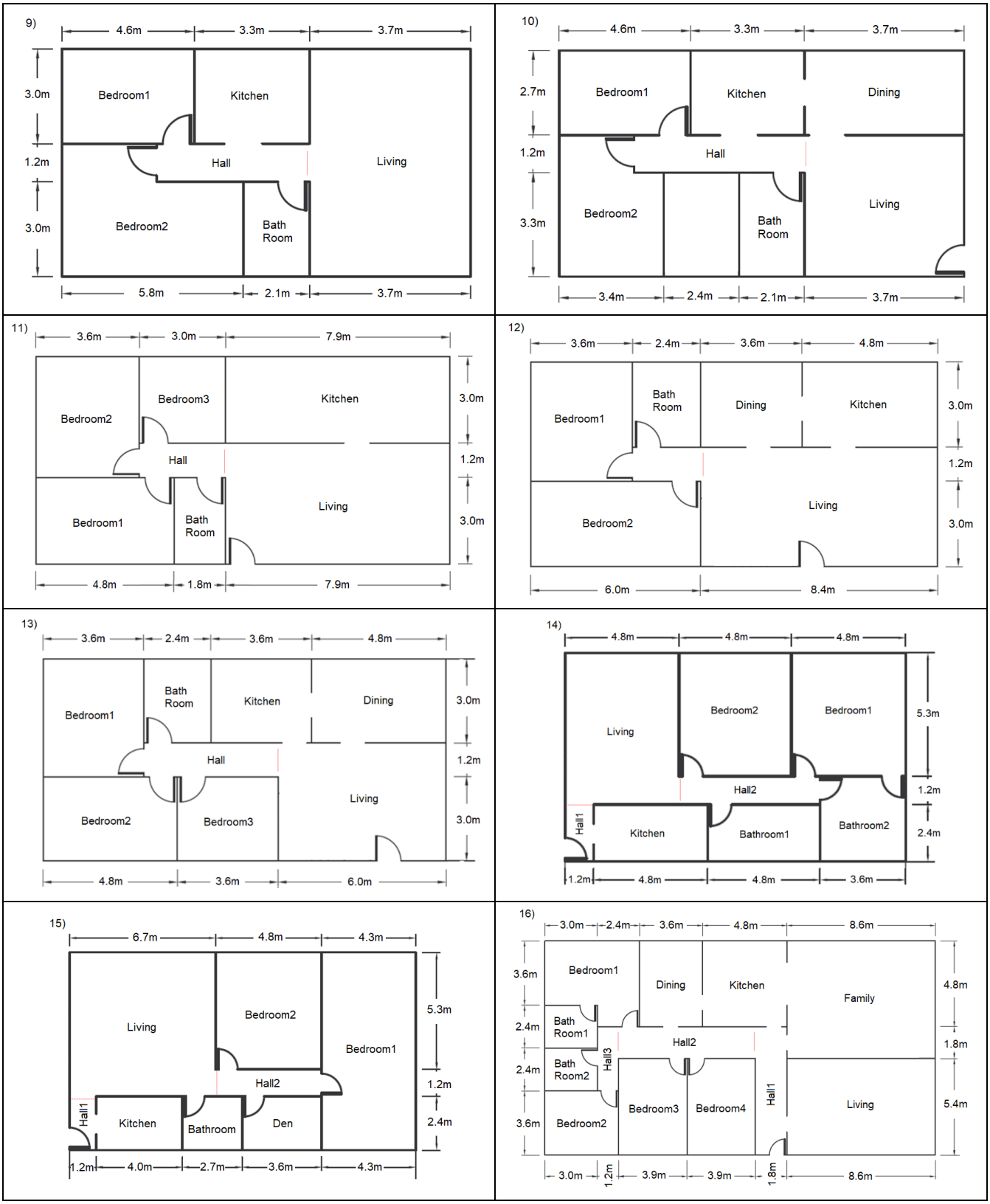


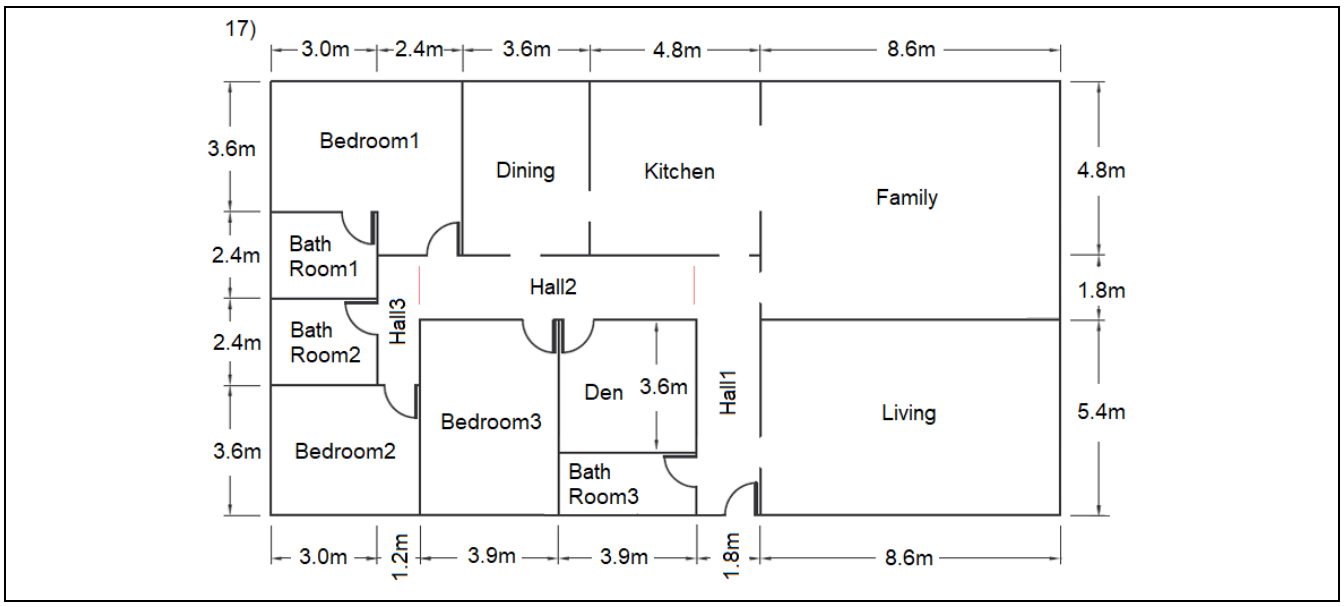
Fig. A1. CFAST validation against measurement for a) Experiment 1 and b) Experiment 2.

Appendix B: Table of Parameters Used for Data Collection

	Parameters	Lower	Upper	Distribution Function	Value/Mean	Variance
Fire	**Flaming Smoldering HRR (Q_1)**	10	30	Uniform	--	--
	Flaming Smoldering Time (t_1)	150	1200	Uniform	--	--
	T-squared Growth HRR (Q_{max})	130	4620	Log-Normal	3250	1000
	T-squared Growth HRR (t_2)	50	330	Uniform		
	Time for Peak HRR (t_3)	--	--	Constant	1000	--
	Decay HRR	--	--	Constant	0	--
	Decay Time (t_4)	--	--	Constant	200	--
Vents	**Window Opening Fraction**	0	1	Binary	--	--
	Window Breakage Threshold	--	--	Constant	150	--
	Front Door Opening Fraction	0	1	Binary	--	--
	Front Door Opening Time	--	--	Constant	0	--
	Front Door Opening Probability	--	--	Constant	50	--

Appendix C: Overall Floorplans of the Seventeen Homes


9)
4.6m
3.3m
3.7m
3.0m
1.2m
3.0m
Bedroom1
Kitchen
Hall
Living
Bedroom2
Bath Room
5.8m
2.1m
3.7m
10)
4.6m
3.3m
3.7m
2.7m
1.2m
3.3m
Bedroom1
Kitchen
Dining
Hall
Bedroom2
Bath Room
Living
3.4m
2.4m
2.1m
3.7m
11)
3.6m
3.0m
7.9m
Bedroom2
Bedroom3
Kitchen
3.0m
Hall
1.2m
Bedroom1
Bath Room
Living
3.0m
4.8m
1.8m
7.9m
12)
3.6m
2.4m
3.6m
4.8m
Bedroom1
Bath Room
Dining
Kitchen
3.0m
1.2m
Bedroom2
Living
3.0m
6.0m
8.4m
13)
3.6m
2.4m
3.6m
4.8m
Bedroom1
Bath Room
Kitchen
Dining
3.0m
Hall
1.2m
Bedroom2
Bedroom3
Living
3.0m
4.8m
3.6m
6.0m
14)
4.8m
4.8m
4.8m
Living
Bedroom2
Bedroom1
5.3m
Hall2
1.2m
Hall1
Kitchen
Bathroom1
Bathroom2
2.4m
1.2m
4.8m
4.8m
3.6m
15)
6.7m
4.8m
4.3m
Living
Bedroom2
Bedroom1
5.3m
Hall2
1.2m
Hall1
Kitchen
Bathroom
Den
2.4m
1.2m
4.0m
2.7m
3.6m
4.3m
16)
3.0m
2.4m
3.6m
4.8m
8.6m
3.6m
Bedroom1
Dining
Kitchen
Family
4.8m
2.4m
Bath Room1
Hall2
1.8m
2.4m
Bath Room2
Hall3
Hall1
Bedroom3
Bedroom4
Living
5.4m
3.6m
Bedroom2
3.0m
1.2m
3.9m
3.9m
1.8m
8.6m



It should be noted that the red lines are for visualization to identify the hallways in different building structures.

Appendix D: Model Performance for Different Neural Network Structures

Table D1: Model performance for different two-layer graph neural network structures.

T-Conv	S-Conv	Accuracy	Precision	Recall	F1
1	4	63.4%	60.1%	79.8%	68.6%
1	8	68.3%	67.8%	69.5%	68.7%
1	16	82.2%	81.0%	84.2%	82.5%
1	32	78.9%	82.7%	73.1%	77.6%
1	64	81.6%	86.8%	74.5%	80.2%

Note that the first layer is the temporal convolutional layer (T-Conv) and the second layer is the spatial convolutional layer (S-Conv).

Table D2: Model performance for different three-layer graph neural network structures.

T-Conv	S-Conv	T-Conv	Accuracy	Precision	Recall	F1
1	4	1	57.3%	57.4%	56.1%	56.8%
1	4	4	64.0%	65.0%	60.6%	62.7%
1	4	8	71.5%	72.3%	69.6%	70.9%
1	4	16	75.9%	72.5%	83.6%	77.6%
1	4	32	83.5%	82.4%	85.1%	83.8%
1	4	64	83.5%	82.3%	85.3%	83.8%
1	8	1	50.1%	57.9%	0.0%	0.1%
1	8	4	64.5%	61.6%	77.0%	68.5%
1	8	8	77.4%	78.6%	75.4%	76.9%
1	8	16	77.0%	72.8%	86.3%	79.0%
1	8	32	80.1%	80.6%	79.4%	80.0%
1	8	64	85.2%	82.7%	89.0%	85.7%
1	16	1	57.1%	57.2%	56.3%	56.7%
1	16	4	71.8%	72.4%	70.5%	71.5%
1	16	8	78.9%	79.0%	78.9%	78.9%
1	16	16	84.8%	84.5%	85.3%	84.9%
1	16	32	80.7%	85.5%	74.0%	79.3%
1	16	64	86.5%	88.1%	84.5%	86.2%
1	32	1	57.7%	55.7%	74.9%	63.9%
1	32	4	74.5%	78.1%	68.2%	72.8%
1	32	8	81.1%	77.2%	88.4%	82.4%
1	32	16	85.2%	83.5%	87.8%	85.6%
1	32	32	85.1%	83.9%	86.9%	85.4%
1	32	64	86.8%	85.8%	88.2%	87.0%
1	64	1	57.8%	59.0%	51.3%	54.9%
1	64	4	77.8%	79.2%	75.6%	77.3%
1	64	8	82.2%	78.7%	88.3%	83.2%
1	64	16	85.7%	85.8%	85.7%	85.7%
1	64	32	85.8%	8.4%	87.8%	86.1%
1	64	64	85.9%	84.0%	88.7%	86.3%

Note that the first, second, and third layers are the temporal, spatial, and temporal convolutional layer, respectively.