https://doi.org/10.1016/j.eswa.2023.119899

Real-Time Flashover Prediction Model for Multi-Compartment Building Structures Using Attention Based Recurrent Neural Networks

Wai Cheong Tam[1,*,‡], Eugene Yujun Fu[2,*], Jiajia Li[1], Richard Peacock[1], Paul Reneke[1], Grace Ngai[3], Hong Va Leong[3], Thomas Cleary[1], Michael Xuelin Huang[4]

[1] National Institute of Standards and Technology, Gaithersburg, Maryland, USA
{waicheong.tam‡, jiajia.li, richard.peacock, paul.reneke, thomas.cleary}@nist.gov

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

[3] Department of Computing, The Hong Kong Polytechnic University, Hong Kong
grace.ngai@polyu.edu.hk, cshleong@comp.polyu.edu.hk

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

Abstract

This paper presents the development of an attention based bi-directional gated recurrent unit model, P-Flashv2, for the prediction of potential occurrence of flashover in a traditional 111 m^2 single story ranch-style family home. Synthetic temperature data for more than 110 000 fire cases with a wide range of fire and vent opening conditions are collected. Temperature limit to heat detectors is applied to mimic the loss of temperature data in real fire scenarios. P-Flashv2 is shown to be able to make predictions with a maximum lead time of 60 s and its performance is benchmarked against eight different model architectures. Results show that P-Flashv2 has an overall accuracy of ~ 87.7 % and ~ 89.5% for flashover predictions with a lead time setting of 30 s and 60 s, respectively. Additional model testing is conducted to assess P-Flashv2 prediction capability in real fire scenarios. Evaluating the model again with full-scale experimental data, P-Flashv2 has an overall prediction accuracy of ~ 82.7 % and ~ 85.6 % for cases with the lead time of setting 30 s and 60 s, respectively. Results from this study show that the proposed machine learning based model, P-Flashv2, can be used to facilitate data-driven fire fighting and reduce fire fighter deaths and injuries.

Keywords: Flashover occurrence, machine learning, real-time prediction, realistic fire and opening conditions, benchmark models.

1. Introduction

Over the past 40 years, many important advances have been made both in understanding the effect of various factors influencing the onset of flashover [1-5] and the development of scientific approaches to forecast the potential occurrence of flashover within residential building structures based on image data [6-9] and time-series data [10-18]. In [6], Yun and his co-workers applied generative adversarial networks to enhance the IR images from the videos to make prediction of

* These two authors contributed equally.
‡ Corresponding author.

flashover based on the dynamic changes of the fire and smoke patterns. Wang et al. [7] developed a convolutional neural network based prediction model with smoke images generated from a fire simulation program. The model could determine the transient heat release rate in the building fire and make used of the estimated heat release rate to further classify flashover conditions. Mozaffari et al. [8] utilized RGB images obtained from typical video cameras to develop an optimized convolutional neural network model to detect flashover occurrence in compartment fires. Li et al. [9] proposed a new hybrid model to predict flashover occurrence for as early as 51 s before it happens. Indeed, the computer vision based approaches from [6-9] are quite promising in terms of prediction capability. However, there are limitations for these models. Firstly, they require the use of cameras that are located either near the floor surface in the fire of origin or being placed outside of the fire room. Secondly, in order to effectively learn the indicative patterns from images to flashover conditions, the size of these models is generally very big. For that, these approaches likely require a lot of computing resources which make the use of these approaches difficult in real world applications. Practically efficient methods [10-14], such as the use of empirical correlations developed based on full-scale experiments, are available to estimate the onset of flashover in single compartment structures with a door-like opening for a wide range of combustible fuels. The effect of multi-compartment structures is also included in many studies [15-18]. However, few of these research advances have been utilized in any significant degree by the fire protection community, particularly in the area of smart fire fighting, where reliable real-time forecast of the potential flashover occurrence is known to be not only important, but life-saving for line-of-duty fire fighters to avoid extreme fire events (i.e., flashover) on the fire ground [19].

Limitations exist in currently available flashover prediction models. Experimental studies [10-14] provided empirical correlations for various aspects, including burning characteristics for different combustibles, the effect of fire locations within a compartment, ventilation, ceiling height, and compartment surface materials, to flashover potential. Although this kind of analytical approach can readily be used in practical fire applications, one primary drawback is that accurate knowledge about the fire and vent opening conditions is needed. Research efforts from [20,21] help to eliminate the dependence of prior knowledge about the fire and vent opening conditions for flashover predictions. In these studies, temperature data were continuously being collected from thermocouple trees when a fire event was detected. Utilizing the sensor data as inputs, hybrid models, which were developed based on the combined use of a fire simulation program [22] and inverse modeling techniques, recover the heat release rate (HRR) and the fire location. Based on this information, the determination of potential flashover occurrence can be made. Yet, these models are limited for use in single compartment structures with either no vent openings or only one single vent opening. Davis et al. [15], Price [16], Koo et al. [17], and Jahn [18] adopt similar approaches from [20,21] and expand the modeling capabilities such that information about the location, HRR, and the spread rate of fire can be retrieved in multi-compartment building structures. In [18], predictions with lead times were also demonstrated to be achievable. These research outcomes are encouraging, but these models require high performance computing (HPC). As mentioned in [18], even using HPC, a single calculation would take approximately 5 minutes to complete. In general, the availability of HPC is rare and/or cannot be precisely allocated during an emergency. Due to time constraints and limited computational resources, these models are unlikely to be suitable for actual fire fighting applications. For that, the development of a more robust model for real-time flashover prediction is vital.

Given a multi-compartment structure, the effectiveness of a prediction model is not only determined based on its numerical efficiency, but also its modeling accuracy. To the best of the authors' knowledge, none of the previous studies [10-21] has attempted to account for realistic conditions that would likely exist in real fire scenarios. In a fire scene, compartment gas temperatures increase significantly. Since typical fire protection sensors, such as heat detectors, are not designed to operate in extreme temperature environments as compared to the measurement devices, such as thermocouples, heat detectors will fail at elevated temperatures. According to NFPA 72 [23], the maximum operational temperature limit for the highest grade heat detectors is approximately 300 °C. When temperature data are lost/unavailable (i.e., gas temperature above 300 °C), the model performance from [15-18] is expected to decrease dramatically. This is because these models require continuous input data. In addition to the sensor limit, the conditions around the vent openings and fire are generally more complicated than those included in [15-18]. It should be noted that these studies assumed vent openings between compartments are always opened and the fire is always located within one compartment of the multi-compartment structure. If the conditions of the vent openings and the fire are different, these models will become highly inaccurate due to the fact that these models were developed based on simplified conditions. When several vents (doors or windows) can be arbitrarily opened or closed, the temperature information between different compartments to the flashover condition becomes more complex. Based on the modeling techniques provided in [15-18], it will be rather impossible for any of these models to provide predictions of potential flashover occurrence in the multi-compartment structures accurately and efficiently. Therefore, the objective of this work is to demonstrate the use of a machine learning paradigm to develop a robust flashover prediction model that can be used in real-time within a multi-compartment residential structure while accounting for realistic conditions, including a heat detector maximum operational temperature limit, and arbitrary vent opening and fire conditions.

In this paper, the development of a recurrent neural network (RNN) based flashover prediction model (P-Flashv2[1]) is presented. In Section 2, the use of the learning-by-synthesis approach is demonstrated for data generation. In this same section, a model validation is provided to ensure synthetic data quality. The details for the model structure of P-Flashv2, as well as the underlying principle for the model formulation are presented in Section 3. In Section 4, results are presented. They include comparison of model performance and testing against full-scale experimental data. Finally, concluding remarks and future work are provided in Section 5.

The contributions of this work are two-fold. This present work aims 1) to provide a comprehensive synthetic data set for the development of a machine learning based flashover prediction model in a multi-compartment structure with a wide range of fire and vent opening conditions, and 2) to establish benchmark model comparison and model testing. It is believed that the output from this work will provide a step forward to reduce fire fighter deaths and injuries[2].

2. Multi-Compartment Fire Data

[1] P-Flash is a machine learning model developed based on recurrent neural network achitectures.

[2] Over the past ten years, nearly 750 fire fighters were killed and approximately 250 000 fire fighters were injured on the fireground [20,21]. Rapid fire progression, such as flashover, has been identified as one of the leading causes for both fire fighter fatalities and injuries.

To develop a robust machine learning (ML) based flashover prediction model, the training dataset must contain a large number of fire experiments covering a wide range of fire conditions with various geometric settings. Due to the fact that flashover is an extreme fire event, the collection of data becomes a challenging task. Unlike the classification tasks [26,27] and/or the regression tasks [28,29] being well established in the AI/ML community, the study of the prediction of potential flashover occurrence in full-scale building structures has not been adequately addressed. For this reason, flashover related data in buildings is not available in any public data repositories such as [30]. In principle, fire data can be obtained by physically conducting the required full-scale experiments. However, a recent study [31] reveals that approximately 1000 cases are needed to achieve model convergence in an idealized three-compartment building structure with simplified vent opening conditions. It can easily be understood that if a more complex building structure (i.e., more compartments) and the realistic conditions (i.e., different fire and vent opening conditions) have to be accounted for, the required number of physical experiments increases dramatically. When factors, such as cost and time, are of concern, physically conducting all full-scale experiments is not feasible. Therefore, in order to facilitate the development of a ML based flashover prediction model, the learning-by-synthesis approach is utilized.

2.1. Data Collection

The learning-by-synthesis (LBS) approach [32] provides an alternative means for data collection. To relate this concept to fire research, the approach promotes the use of computer simulation programs to carry out the desired full-scale experiments numerically. Since the cost and time associated with numerical experiments are considerably lower than physically conducting the full-scale experiments, LBS provides flexibility to optimize the data quality through trial-and-error and to ensure the acquisition of a large amount of relevant data.

In the fire research community, LBS plays a significant role in the data collection process in recent studies. In [31], a zone model is used as the simulation engine to facilitate data generation for various fire scenarios within a simple building structure. Other studies, such as [33], attempt to use a field model [34] to obtain the required fire data in a tunnel with different fire and ventilation settings. In the present study, the zone model, namely CFAST [22], is used. CFAST is a two-zone fire model that predicts the thermal environment caused by a fire within a compartmented structure. Each compartment is divided into an upper and lower gas layer. The fire drives combustion products from the lower to the upper layer via the plume. The temperature within each layer is uniform, and its evolution in time is described by a set of ordinary differential equations derived from simplified mass and energy conservation. The transport of smoke and heat from one zone to another is governed by empirical correlations. Because of the zone assumption, the modeling of fluid flow within the computational domain is not performed. For this reason, if the effect from the spread of hot gas becomes significant and the zone assumption becomes invalid (i.e., a warehouse like structure with a large floor area or a closet like compartment with an extremely small floor area), the use of a more sophisticated fire simulation program, such as [34], is more appropriate.

Certainly, two questions can be raised for the use of CFAST: 1) does the model have the modeling capabilities to produce reliable synthetic data for the domain of interest in the current problem settings, and 2) since the model is being trained based on CFAST data, can the model be trusted to be used in real fire scenarios. In order to address the first and the second questions, model

validation for CFAST (refer to Sec. 2.1.2) and model testing of the machine learning based model against full-scale experimental data (refer to Sec 4.2) are conducted, respectively.

Even though the concept of LBS seems a promising solution to overcome the challenge associated with data collection, an additional maneuver is needed because manually creating CFAST input files is a time consuming task. In order to carry out the data collection process systematically, CData is utilized to automatically generate the desired number of CFAST input files. As described in [31], CData is an in-house program that can operate on a Linux cluster. Given the user-specified simulation parameters, such as building layouts, surface materials, fire conditions, ventilation configurations, location of detector(s), and output intervals, the program provides corresponding CFAST input files based on seven different probability density functions. The user-defined functions can formulate inputs for desired fire environments. Depending on the problem complexity and computational resources of the Linux cluster, approximately 10 000 CFAST cases can be executed and completed in a day. A summary spreadsheet containing detailed information about all input parameters for each of the cases is also generated. This spreadsheet can be used to carry out data inspection to eliminate any duplicate cases and to examine the data behavior of each of the simulation parameters. This information is crucial to ensure the quality and the covering range of the fire data. A detailed discussion of the numerical setup for the present study is elaborated in the following sections.

2.1.1. Numerical Setup

Single Story Structure

Consider a single story ranch-style structure as shown in Fig. 1a. There are seven major compartments: a living room, a dining room, a kitchen, three bedrooms, and a hallway connecting the living room and the three bedrooms. The overall approximate interior dimensions of the structure are 13.92 m x 7.7 m with a ceiling height of 2.44 m. The detailed dimensions associated with each of the compartments are shown in Fig. 1a. Since fire rarely occurs in bathrooms, they are not considered in the current layout. For the interior finish, the walls, ceiling, and floor of each compartment are constructed with gypsum wallboards. In some occasions, concrete is also used as building material. In order to examine the effect associated with different building materials, numerical experiments are carried out to obtain data for two different building materials. The approximate thermal properties of the two materials are summarized in Table 1.

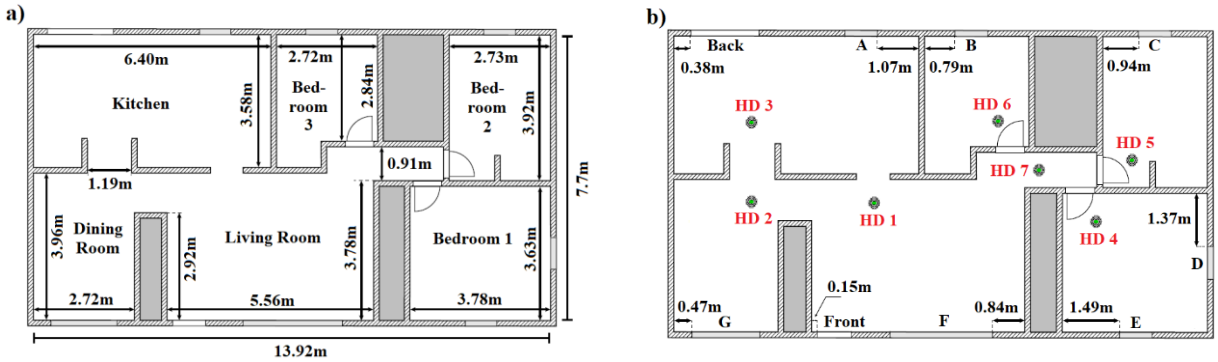


Figure 1. Plan view dimensioned drawing of a) the single story structure [36] and b) vent openings with heat detectors (HD). Note that the heat detectors are added to the simulations.

Table 1. Summary of approximate thermal properties for building materials.

Materials	**Conductivity** kW/(m·K)	**Specific Heat** kJ/(kg·K)	**Density** kg/m^3	**Thickness** m	**Emissivity**
Gypsum	0.00016	1	480	0.025	0.9
Concrete	0.0016	0.75	2400	0.15	0.94
Glazing	0.0008	0.8	2500	0.003	0.95

Figure 1b shows the relative position of the vents and the heat detectors in different compartments. For vents, there are two exterior doors (front and back), three bedrooms doors, and seven windows (A through G). The vent sizes and sill heights are summarized in Table 2. One heat detector is located at each compartment, and they are approximately 0.02 m down from the ceiling. The response time index for the heat detector is 35 $(m\text{-}s)^{0.5}$. It is worth noting that this single story, traditional ranch-style structure is selected because about 90 % of residential buildings were built using this layout since the mid-1950s [36]. Currently, this structure remains the most popular style of home in 34 states across the United States [37]. The flashover prediction model developed based on this structure is expected to have substantial benefits for fire fighting across the U.S. Also, another main reason for the selection of this building layout is that full-scale experimental data are available for model verification and model testing. Additional efforts accounting for the effect associated with different structure layouts is underway. Findings will be reported in future studies.

Table 2. Summary of approximate vent size and sill height [32].

Vents	**Size**	**Sill Height**
Front	0.89 m x 2.05 m	0 m
Back	1.78 m x 2.05 m	0 m
A	0.85 m x 1.02 m	1.07 m
B	0.86 m x 1.46 m	0.61 m
C	0.86 m x 1.46 m	0.61 m
D	0.86 m x 1.46 m	0.61 m
E	0.85 m x 1.46 m	0.61 m
F	2.67 m x 1.46 m	0.61 m
G	1.78 m x 1.46 m	0.61 m

Fire and Vent Opening Conditions

For fire and vent opening conditions, three realistic conditions are taken into account in this study, and they are 1) the use of experimentally validated fire growth of single burning items, 2) consideration of various fire locations, and 3) arbitrary opening conditions of vents.

Realistic Fire Growth of a Burning Item: HRR is the single most important variable in characterizing the fire growth of an item [38]. In order to obtain the HRR, experiments are typically performed under a furniture calorimeter, and Figure 2 shows the standard HRR curve of a burning item. Specifically, the fire growth behavior is described by a t-squared power function with respect to time. Although more complex burning behaviors (i.e., a fire growth curve with multiple peaks) are often found in typical items, Babrauskas and his co-worker [5] demonstrated that the details of the shape of the fire growth curve up to the time of flashover are of secondary importance. They

further commented that even curves with multiple HRR peaks prior to flashover have little impact on the HRR value found at flashover. As it was shown in their study, given enough oxygen, the primary factor causing the onset of flashover is the HRR value and how fast an item would cause flashover depends on its HRR/time relation. For this reason, the current study adopts the t-squared fire growth curve to describe the burning behavior of a single item. It should be noted that this burning behavior has also been well observed in many classic experimental studies [39,40].

Similar to the plot shown in Fig. 2, a burning item typically goes through four different fire growth stages: linear growth (i.e., smoldering fire), t-squared growth (i.e., flaming fire), peak, and decay. In the current study, three items are considered, and they are flaming chairs, polyurethane foam mattresses, and cotton based mattresses. The transition HRR from smoldering to flaming fire (Q_1), peak HRR (Q_{max}), time to transition (t_1), time to peak HRR (t_2), peak time ($t_3 - t_3$), and decay time ($t_4 - t_2$) are summarized in Table 3. The peak HHR and time to peak HRR are obtained from [41], and the fire growth rate, α, is determined between 0.000329 kW/s^2 and 0.041387 kW/s^2, ranging from slow to fast fire growth rate. In general, the selection of these items follows the primary burning items in [36]. The transition HRR and time to transition HRR are selected based on examination of HRR data provided in [42]. Since the flashover usually occurs during the t-squared growth stage and peak stage, the exact value for peak time and decay time is not significant. Also, the incipient stage of the fire is highly variable, as opposed to the t-squared flaming growth stage.

Probability density functions (PDF) are used to facilitate case samplings and the use of a specific PDF is based on numerical experiments such that a wide range of fire growth behavior is included and sufficient data for flashover cases are obtained. A uniform PDF is used to sample peak HRR for the three different items and a skew normal PDF is used to sample time to peak HRR. This is because a ML based model is intended to be developed for potential flashover prediction and there is a significant data imbalance associated with non-flashover related and flashover related data. Also, the motivation of using different PDFs is to include more flashover cases (i.e., more intensive rapidly growing fires) in the data set so that the ML based model can learn the important relationship about potential flashover conditions. For transition HRR, time to transition, peak time, and decay time, a uniform probability density function is utilized.

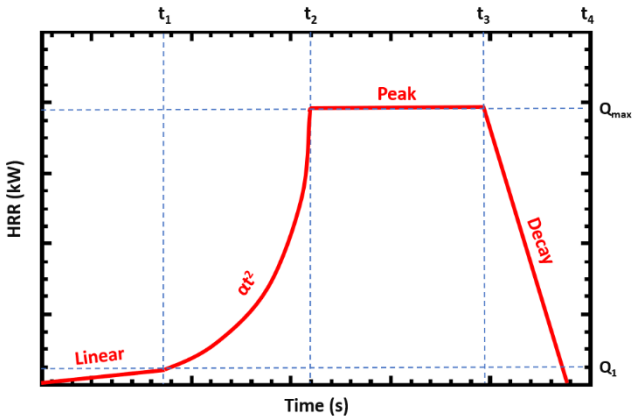


Figure 2. Overview of HRR in different fire growth stages for a typical t-squared fire.

Table 3. Approximate HRR parameters for chair and mattresses [41, 42].

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	200	10
Mattress (foam)	10 – 30	2275 – 4620	150 – 1200	305 – 435	200	10
Mattress (cotton)	10 – 30	130 – 1670	150 – 1200	360 – 1240	200	10

Fire Location and Vent Conditions: A fire can be initiated on the floor at the center, against a wall surface, or at a corner within a compartment. It can also take place in any one of the six different compartments as shown in Fig. 1a. Although a fire can occur anywhere within a compartment, the exact location of the fire is generally not critical due to the assumption inherent in a zone model. However, fire against a wall surface or at a corner leads to an enhanced effect to the HRR due to radiation feedback from the wall surface and/or reduced entrainment due to the presence of wall surfaces. In CFAST [22], the HRR will be double or quadruple for a fire located against a wall surface or at a corner, respectively. In addition to the horizontal location of a fire, the vertical location of the base of the fire plays an important role in the fire growth. Since the hot gas layer is usually filled with by-products of combustion, the base of the fire is always located on the floor surface. This setting allows entrainment of oxygen-rich air and facilitates a more idealized condition for burning. In contrast, the fire growth of a burning item is expected to be diminished due to a lack of oxygen at an elevated location. For that, the base of the fire in this study is always located on the floor surface. For sampling, a uniform PDF is utilized to generate the identical number of cases for fires at three different locations in six different compartments (except the hallway).

For vent opening conditions, all doors and windows are initially closed and they can be opened based on the following two settings: a time-trigger setting for doors and a temperature-trigger setting for windows. For the time-trigger setting, doors in any compartments within the building structure can be opened at any given time within the time to peak HRR (t_2) as shown in Fig. 2. A skew normal PDF is used and each door will have a 50 % chance to be opened. This setting is expected to mimic door opening events due to activities such as leaving the room of the fire origin and/or leaving the building structure. Also, the arbitrarily open conditions of doors indirectly introduce more complex fire growth behaviors to the numerical experiments. For example, assuming that the door and the two windows in Bedroom 1 as shown in Fig. 1a are initially closed, if a fire occurs in Bedroom 1, there is not enough oxygen to sustain the fire growth to reach flashover conditions. However, if the door is opened due events such as evacuation, fresh air will enter the room and the fire will continue to grow to its peak if oxygen is sufficient. In these situations, there are two fire growth scenarios and the fire data (i.e., temperature or species concentration) are expected to be significantly different. When all vents are closed in Bedroom 1 (denoted as Scenario A), the temperature will continue to increase. But, due to lack of oxygen, the temperature is likely to level off and eventually drop. For that, flashover conditions will not be met. For Scenario B where the door is opened, the temperature behavior will change and there will be an increase in temperature. This kind of double-peak fire growth behavior is not captured in a simple t-squared fire. If the fire is large enough, flashover conditions might be met. It is believed

that both scenarios do occur in real structure fires and the arbitrarily opening of doors helps to generate numerical experiments to obtain these data.

The temperature-trigger setting allows windows to be arbitrarily opened when a temperature threshold is reached. In a fire scenario, a window is subjected to heating either from the local gas temperature and/or from the fire. Due to temperature or a flux gradient [43], a window may crack and eventually break out and create an opening to the outside environment. There have been studies that quantify the range of lower bound gas temperatures for window breakage. In this study, single-pane windows (ordinary float glass) with thickness varying from 3 mm to 6 mm are considered. Based on [44], breakage of a single-pane float glass is experimentally observed at temperature between 100 °C and 200 °C. 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 [44]. The thermal properties of the glazing are summarized in Table 1. In general, when the temperature of a target reaches a threshold, the corresponding window will then be opened. A uniform PDF is utilized to sample the window breakage temperature threshold. Similar to that of a door opening, arbitrarily opening windows facilitates the generation of realistic conditions in fire scenarios, and it provides more complex data behavior for development of a more robust ML-based model. It is worth noting that although only one building layout is considered, the data set needed to cover the realistic conditions about fire and vent opening conditions is enormous.

2.1.2. 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 [45], where a fire initiated the living room within the single story residential structure is used to benchmark the synthetic data. The fire location and the HRR of the gas burner fires are the same for the two tests (Exp 1 and Exp 2). Yet, opening conditions of each of the vents are different. The details of the opening time for each vent is provided in Table 4. 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.

Table 4. Event sequence.

Event	Exp 1	Exp 2
Front Door Open	300 s	1200 s
Back Door Open	1275 s	1860 s
Window A Open	1260 s	1845 s
Window B Open	1245 s	1830 s
Window C Open	1230 s	900 s
Window D Open	1215 s	Closed
Window E Open	600 s	600 s
Window F Open	1200 s	1815 s

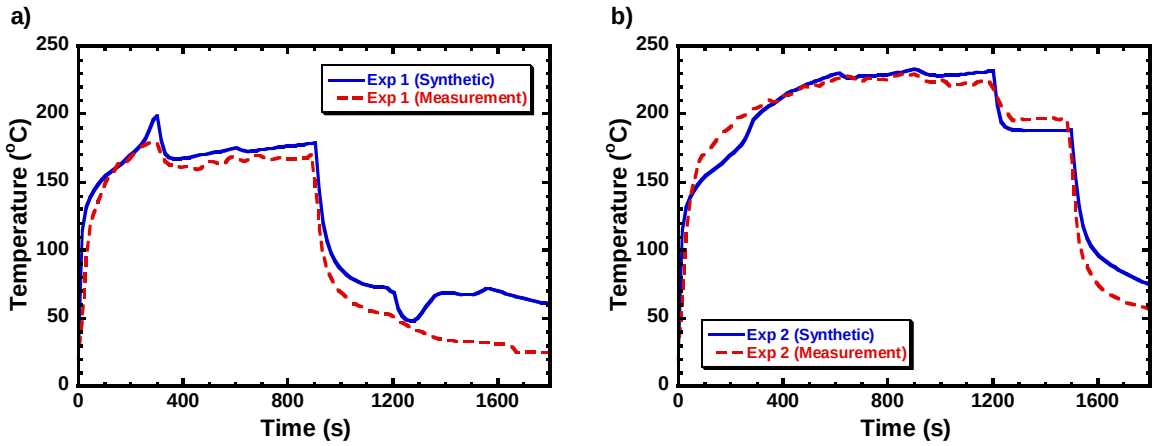


Figure 3. CFAST validation against measurement for a) Experiment 1 and b) Experiment 2.

Figures 3 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, and they are obtained based on the hot gas layer reduction method provided in [30]. 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. Comparison is also made for 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. It is worth noting that CFAST is validated against more than 15 other sets of full-scale experiments with peak heat release rate (HRR), compartment aspect ratio (i.e., compartment length against 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. Validating against these experimental data, CFAST predictions of upper layer gas temperature average within 6 % of experimental measurements [46].

2.1.3. Temperature Data Behaviors and Potential Modeling Challenges

Understanding the temperature data behaviors is vital to the design of a robust model architecture such that the ML-based prediction model is capable of learning the important relationship between temperature and flashover conditions. In this section, two different fire scenarios are highlighted. Fig. 4a shows seven heat detector temperatures and a temperature indicator for a medium growth fire with high peak HRR occurring in the living room. In this case, the front door is open and all interior doors are closed. Given sufficient oxygen, the fire continues to grow and the detector temperature increases accordingly until the fire goes out. It is worth noting that the temperature indicator (dash line) is the upper layer gas temperature for the room of fire origin, and it is used to determine when the temperature threshold for flashover condition is met.

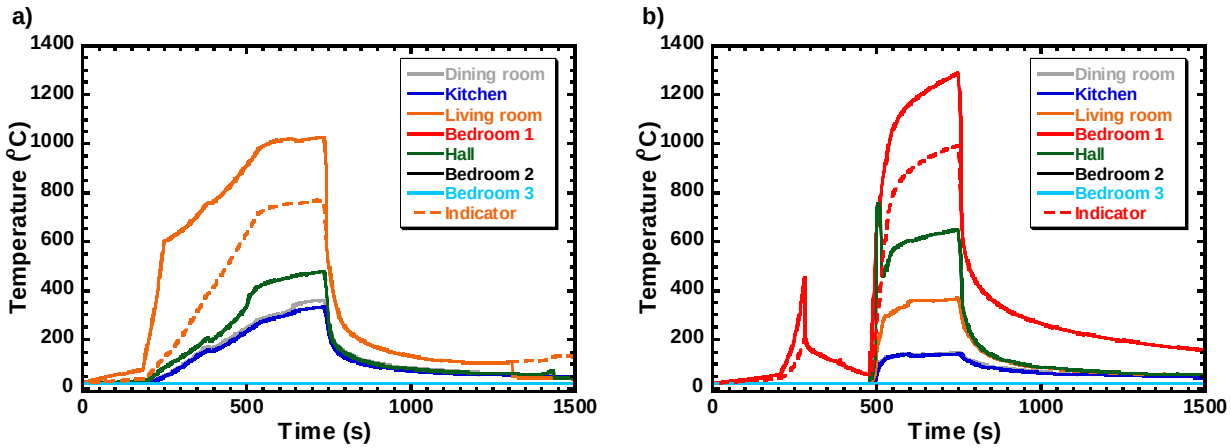


Figure 4. Temperature profiles of different compartments for a) fuel-controlled fire case and b) ventilation-controlled fire case.

Fig. 4b shows temperature data for a more complex fire situation that is similar to that described in Scenario B in section 2.1.1. A fire occurs in Bedroom 1. All the doors and windows in the building structure are closed. This setting initially creates a "sealed" compartment for Bedroom 1. As seen in the figure, the temperature first increases and then decreases due to the lack of oxygen. At around 470 s, the door is opened and the fire is approaching its peak. The temperature increases in bedroom 1, living room, kitchen, and dining room, but the temperatures in Bedroom 2 and 3 remain relatively close to initial temperature. Physically, it is completely understandable because the doors are closed and the heated gases cannot enter the bedrooms. However, the information about the opening of the doors in not known in a real fire scenario. For that, the model needs to have capabilities to determine the temperature data with higher importance.

2.1.4. Defining Thresholds

Heat Detector Maximum Operational Temperature Limits

Loss of heat detector temperature signal is another realistic condition that is being accounted for in this study. In actual fire scenarios, heat detectors are very unlikely to survive at elevated temperature [47] and would fail at temperatures well below the flashover conditions. According to NFPA 72 [48], heat sensing fire detectors are categorized into seven different classes with temperature classifications ranging from low to ultra-high, and the maximum operational temperature ranging from approximately 29 °C to 302 °C. Table 4 provides the temperature classification for the different heat detectors. As an illustration, Fig. 5 depicts the temperature profiles as shown in Fig. 4b with a temperature cut-off at 150 °C. As seen in Fig. 5, the lower the maximum operational temperature of the heat detectors, less temperature signals are available. With less temperature information, the prediction of the potential flashover occurrence will become more difficult. Therefore, in addition to being able to discriminate temperature data with high importance, the model also needs to have learning capabilities to correlate complex temperature information from other compartments to flashover conditions in the room of fire origin.

Table 4. Temperature classification and approximate temperatures for heat sensing fire detectors.

Temperature Classification	Temperature Range (°C)	Maximum Ceiling Temperature (°C)
Low	39 to 57	28
Ordinary	58 to 79	47
Intermediate	80 to 121	60
High	122 to 162	111
Extra high	163 to 204	152
Very extra high	205 to 259	194
Ultra high	260 to 302	249

Flashover Criteria

Flashover is the near-simultaneous ignition of almost all of the directly exposed combustible material in a compartment. Numerous variables can affect the transition of a compartment fire to full room involvement (i.e., flashover) [4]. Thermal influences including radiative and convective heat flux are assumed to be the driving forces and are clearly important. Ventilation conditions, compartment volume, and the chemistry of the hot gas layer can also influence the occurrence of flashover. Based on the experimental studies reviewed in [4], the onset of flashover within a compartment can be quantified by two measurable criterions: 1) heat flux and 2) temperature. Peacock and his co-workers [4] demonstrated that when the incident heat flux onto the floor surface is between approximately 15 kW/m^2 to 33 kW/m^2, there can be a potential occurrence of flashover. However, the measurement of heat flux can rarely be measured in typical building environments because heat flux gauges are usually not installed. For that, the flashover criteria associated with heat flux cannot be used, and it will not be considered in this study. For the temperature criterion, the range of values obtained from nine different literature sources in [4] ranges from 450 °C to 771 °C. The wide range of values for the temperature criterion is due to the large change in temperature within a very short period of time. Yet, it can be observed from [4] that most of the values are in the 550 °C to 650 °C range and the possible flashover temperature is shown in Fig. 5. To be conservative, the upper gas layer temperature of 600 °C is used as the threshold for the potential occurrence of flashover in this study.

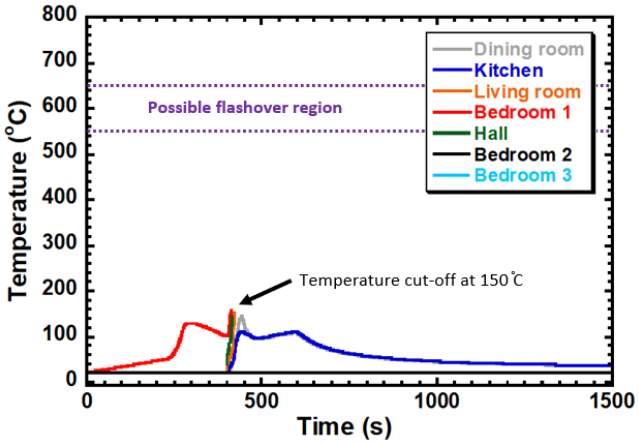


Figure 5. Temperature profiles of different compartments for Scenario B with a heat detector temperature cut-off value of 150 °C.

3. Development of the Prediction of Flashover Neural Network Model (P-Flashv2)

3.1. Recurrent Neural Network

Among many widely used neural networks [49-51], the recurrent neural networks (RNN) [51] have model architectures that are the most suitable for the multivariable time-series classification task which is being considered in this study. Different from the feed-forward neural network [49] and the convolutional neural networks [50], RNN consists of recurrent hidden states [51], and the hidden states allow the temporal dynamic behavior and the sequential information of the data to be both captured and persisted. For the development of the Prediction of Flashover neural network (P-Flashv2), the RNN mode architecture is utilized.

For the development of P-Flashv2 in complex fire situations similar to those shown in Fig. 4b, making use of all available temperature information is crucial. This is because even though the rapid temperature increase in Bedroom 1 appears at different time-stamps (i.e., ~ 250 s and ~ 500 s), the temperature increase does not always lead to a potential occurrence of flashover. In addition, the prediction model will have to neglect temperature data with less significance. For example, since the doors from Bedroom 2 and Bedroom 3 are closed throughout the entire numerical experiment, the corresponding heat detector temperatures from the two bedrooms remained at ambient temperature. In principle, standard RNNs may not have sufficient capabilities to distinguish the physical significance associated with temperature data from different compartments. For that, in order to facilitate the learning of inherent patterns and complex relationships between temperature signals from non-fire compartments and flashover conditions from the room of fire origin in various fire scenarios and with arbitrary vent opening conditions, the bi-directional gated recurrent unit (BiGRU) [52] with attention mechanism [53] is used. In this following subsection, the details of the model structure are presented.

3.2. Model structure

P-Flashv2 will need to be able to carry out the following two tasks: 1) to relate complex data behavior to flashover conditions accounting for the effect of different fire and vent opening conditions and 2) to discriminate data with higher significance and encode contextual information. Fig. 6 depicts the overall model structure of P-Flashv2, consisting of a BiGRU layer, three dense layers, and one attention layer. In general, the model takes only the temperature signals of each of the heat detectors from all compartments, $(\{S^1, S^2 \dots, S^N\}, N = 7)$, as inputs and provides a binary classification, (flashover or non-flashover), as the output. Since the model structure of P-Flashv2 is complex, the descriptions of BiGRU, attention mechanism, and layer configurations are separately given in section 3.2.1, 3.2.2, and 3.2.3, respectively.

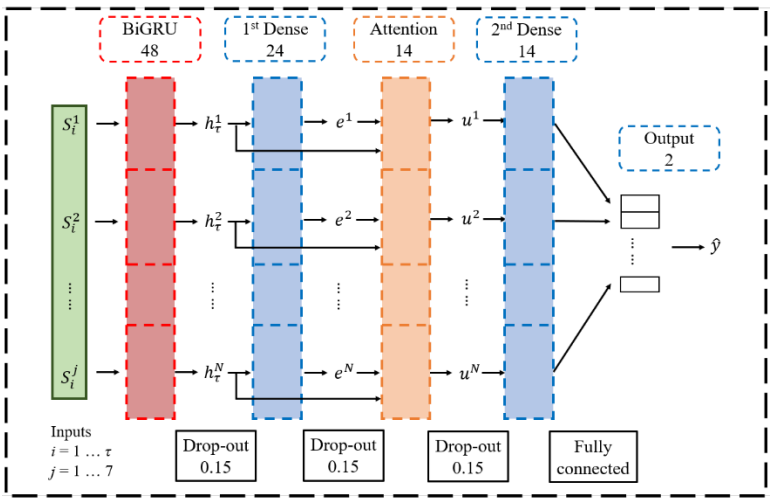


Figure 6. P-Flashv2 model structure with the presence of the BiGRU layer, three dense layers, the attention mechanism, and layer configurations.

3.2.1. Temperature Signal Learning

In order to overcome the challenge associated with the 1st task, the bi-directional gated recurrent unit (BiGRU) is used. Fig. 7a shows the overall model structure of the BiGRU, and Fig. 7b shows the schematic of a single GRU cell from the BiGRU. Within the single GRU cell, there are four primary properties: memory state, hidden state, reset gate, and update gate. At each time step, a memory state is obtained based on two types of inputs: detector temperature data at the current time step (s_i) and hidden states/outputs from the previous time step (h_{i-1}). The amount of information from the previous time step being retained in memory states is regulated based on the reset gate. If the previous hidden state is important for relating the target (i.e., the flashover conditions in this study), the information will be kept. Mathematically, the memory state, c, at time step, i, can be expressed as [48]:

$$c_i = \tanh(W_c[r_i * h_{i-1}, s_i] + b_c) \quad (1)$$

where h_{i-1}, s_i, W_c, b_c, and tanh are the hidden state from the previous time step, temperature data of a heat detector at the current time step, weight and bias for the memory state, and the tangent hypobaric function, respectively. The weight and the bias are the parameters that have to be optimized through the model training process. The parameter, r, is known as the reset gate and it is given as:

$$r_i = \sigma(W_r[h_{i-1}, s_i] + b_r) \quad (2)$$

with σ, W_r, and b_r being the sigmoid function and the weight and bias for the reset gate, respectively. The selection of the tangent hyperbolic function in Eq (1) and the sigmoid function in Eq (2) is based on many in-depth studies such as [50] in which the GRU model structure is optimized for a wide range of problems. From Eq (2), it can be seen that if the reset gate approaches zero, the effect associated with the hidden state from the previous time step becomes negligible.

In general, the value of the memory state ranges from – 1 to 1 and the reset gate value varies from 0 to 1.

Inside the GRU cell, there is another gate, namely the update gate, that controls how much information from the memory state is being accepted to formulate the new hidden state. The update gate, *z*, is expressed as:

$$z_i = \sigma(W_z[h_{i-1}, s_i] + b_z) \tag{3}$$

with W_z, and b_z being the weight and bias for the update gate and the new hidden state is given as:

$$h_i = z_i * c_i + (1 - z_i) * h_{t-1} \tag{4}$$

which is the sum of the hidden state from the previous time step and the memory state at the current time step. It can be seen that if the information from the current memory state is not important for relating the flashover conditions, the update gate will become zero. With that, the new hidden state will basically be the hidden state from the previous time step. The hidden state will only be updated if important information to the target is recognized. Due to the nature of the sigmoid function, the value for the update gate is varied between 0 and 1. It should also be noted that the two gates from the GRU cell help overcome the long time-dependent issues (i.e., gradient explosion where a model is not converging [52]) from which traditional RNNs such as simple RNN [55] suffer.

As shown in Fig. 7a, the BiGRU consists of two GRU units. It can be seen that the temperature signals from a heat detector, $S^j = (s_1, s_2, s_i, \dots, s_\tau)$, are fed into both the forward GRU unit and the backward GRU unit. The forward GRU takes temperature data from time step $i = 1$ to $i = \tau$ where τ is the moment when the flashover condition is met. The first time step does not need to be at the beginning of the numerical experiment and the exact time stamp can be determined based on the desired/required length of the temperature data. Likewise, the backward GRU takes information from $i = \tau$ to $i = 1$. With that, $\overrightarrow{h_i}$ is the forward hidden state and $\overleftarrow{h_i}$ is the backward hidden state. It should be noted that the reason for using BiGRU over the unidirectional GRU is that BiGRU can learn more effectively because it makes use of all information (i.e., information from the past and in the future) to determine the current hidden state whereas unidirectional forward GRU only uses information in the past and important information towards the beginning can be difficult to remain preserved. In this study, since the complete behavior for temperature signals is needed, only the last hidden states of $\overrightarrow{h_\tau}$ and $\overleftarrow{h_\tau}$ are outputted. As shown in Fig. 7a, concatenation is applied to yield $h_\tau = [\overrightarrow{h_\tau}, \overleftarrow{h_\tau}]$ to encode temperature behavior with flashover conditions, and this information will be passed onto the next layer.

The robustness of the bi-directional RNN model structure and the performance of BiGRU is compared to four other RNN variants, such as Basic RNN [55], unidirectional GRU [56], long short-term memory [57], and bi-directional long short-term memory [58], in other literature.

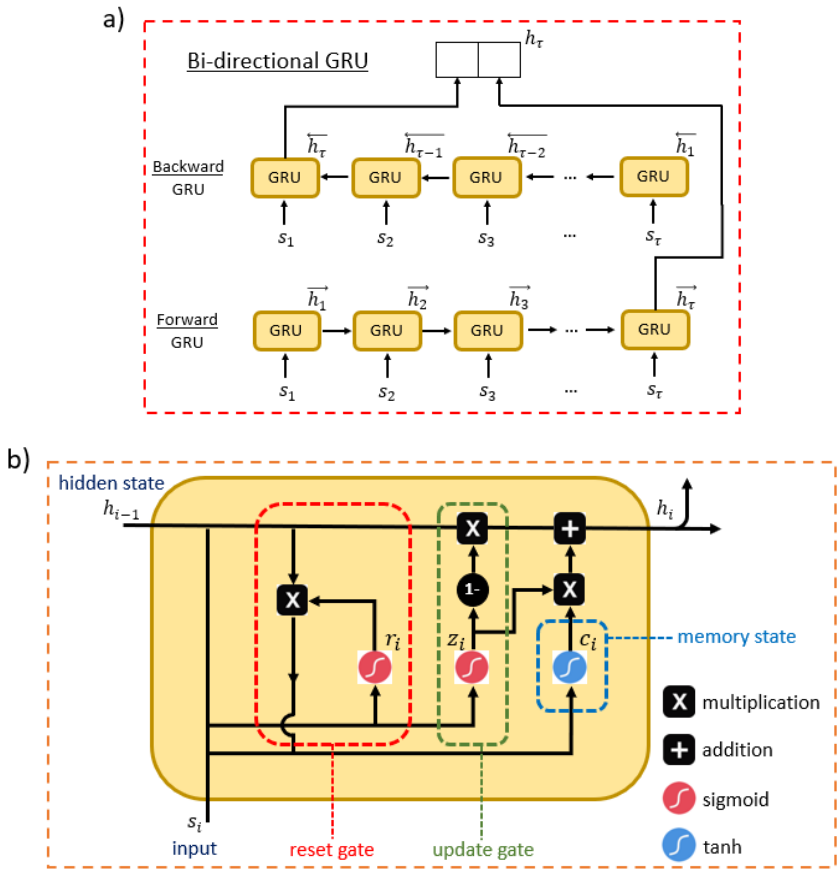


Figure 7. Schematic of a) model structure of BiGRU and b) a single GRU cell.

3.2.2. Sensor-Wise Self-Attention

In order to enhance the learning capability of the model in discriminating temperature signals with higher significance (i.e., neglecting temperature signals from Bedroom 2 and Bedroom 3 in Fig. 4b), a self-attention mechanism is added to correlate sensor-wise relations. It is expected that contextual temperature information of all compartments within the structure can be extracted and the contextual information can contribute to provide more accurate flashover predictions.

As shown in Fig. 6, the temperature signals from all compartments ($\{S^1, S^2, \dots, S^N\}$ for $N = 7$) are the inputs. The hidden state of the temperature behavior (h_τ) for each signal is obtained from the BiGRU, and they are fed into the sensor-wise self-attention module for learning the sensor relation. Mathematically, the attention weight (β_{ij}) of each pair of sensor signals (S^i, S^j) is determined based on the interaction of their temperature behavior (h_τ^i, h_τ^j):

$$\alpha_{ij} = {h_\tau^i}^T h_\tau^j \tag{4a}$$

$$\beta_{ij} = \frac{\exp(\alpha_{ij})}{\sum_{k=1}^{N} \exp(\alpha_{ik})} \tag{4b}$$

To obtain the contextual temperature information that captures the temperature behaviors of all compartments, context features based on the learned attention weights need to be extracted. For a signal S^i, its context feature, g^i, is computed as:

$$e^i = \sigma_{ReLU}(W_e h_\tau^i + b_e) \tag{5a}$$

$$g^i = \sum_{j=1}^{N} \beta_{ij} e^i \tag{5b}$$

where σ_{ReLU} is the rectified linear unit (ReLU) activation function and W_e and b_e are the parameters of the 1st dense layer. As seen in Fig. 6, the size of the outputs from the dense layer is reduced. Instead of allowing all the hidden states to be used in the attention module, a constraint is applied through the use of a dense layer to enhance the selection of better features which will help to improve the model robustness. The exact number of nodes is determined based on numerical experiments, and it will be discussed in section 3.2.3. Another dense layer is applied to concatenate the context features to encode selective feature, $u^i = [g^i, e^i]$, such that the final feature representation v^i for S^i is obtained:

$$v^i = \sigma_{ReLU}(W_c u^i + b_c) \tag{6}$$

The feature representations for all temperature signals, $S^1, S^2, \dots, S^7$, are obtained by the same manner, and the overall representation of the whole structure: $v = [v^1, v^2, \dots, v^7]$ is constructed. This feature representation is used to predict whether there is a flashover occurrence based on the available temperature signals.

In order to assess the effect of the attention mechanism, the attention based RNN model structure is compared to the stand-alone RNN model structure and results will be discussed in Section 4.

3.2.3. Model Configurations

The dimensions of each layer are presented in Fig. 6. In principle, the determination of the number of layers and the corresponding nodes of the specified layer is based on a trial-and-error process, known as fine-tuning [60]. During this process, the model is optimized for prediction accuracy without being either underfitted or overfitted. In general, underfitting refers to the fact that the model is not adequately trained and the model does not learn sufficient relationships to correlate the inputs and the outputs. The latter term refers to over training the model. If the model is overfitted, it tends to memorize the data without learning any generalized information between the inputs and the outputs.

Fig. 8a and 8b show an example of underfitting and overfitting, respectively. Loss (error) and accuracy for both the training data set and the validation data set are plotted against epoch (iteration). The training set is used to train the model and the validation set is to validate the trained model. Note that the model does not see any of the data from the validation set during the model training process.

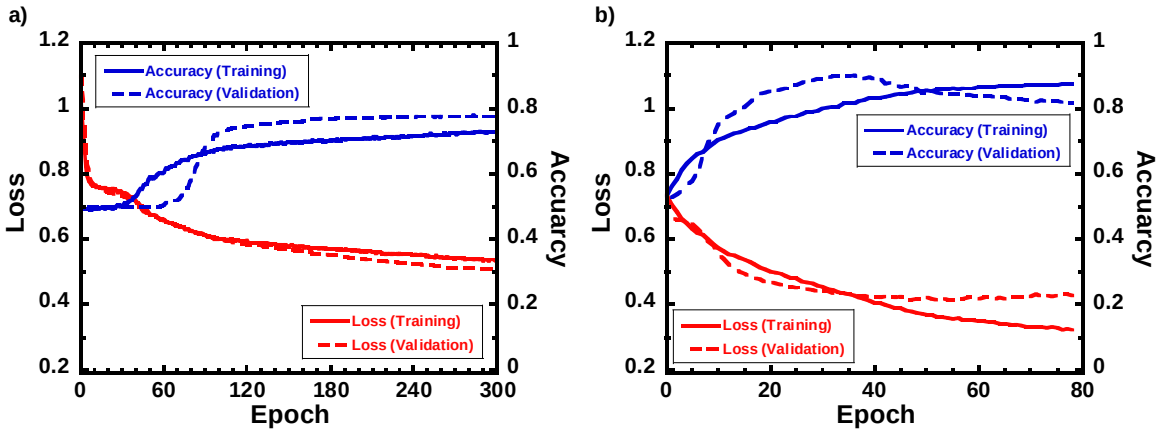


Figure 8: Loss vs. epoch and accuracy vs. epoch plots for training and validation sets for a) underfitting and b) overfitting.

As seen in Fig. 8a, even after 300 epochs, the loss remains high and the accuracy is relatively low for both data sets. In order to resolve the underfitting problem, a more complex model structure (i.e., adding more nodes or even more layers) can be applied such that the learning capability of the model is enhanced. In Fig. 8b, the training loss continues to decrease, but the validation loss stops decreasing at around epoch 50. For accuracy, it gradually increases for the training set. However, the accuracy for the validation set drops at around epoch 35. To avoid overfitting, a common practice is to introduce an early-stop in the model training. That is, if the validation loss does not improve for a given consecutive epochs (i.e., 5 in this study), the training is ended. Another approach is to reduce the model complexity and/or to add more relevant data. It is worth noting that the fine-tuning process for the development of a neural network based model is lengthy.

In the current study, the following model configurations provide the most robust model. The dimension of the BiGRU, 1st dense layer, attention layer, and the 2nd dense layer is 48, 24, 14, and 14, respectively. A dropout rate is set to be 0.15 to help avoid overfitting. The output layer is fully connected, and its dimension is two. Fig. 9 shows the loss and accuracy for the training set and validation set. It can be seen that the validation loss and accuracy approach a converging value for a proper trained model with the right number of nodes and model structure without under- and over-fitting. Early stop [55] with a patience of 10 epochs is used to check if the validation loss does not improve for 10 consecutive epochs.

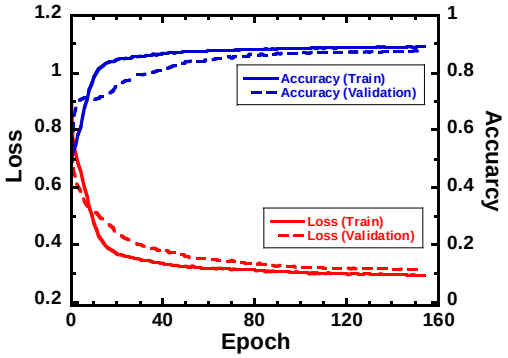


Figure 9. Loss and accuracy plot for optimized P-Flashv2.

3.3. Data Preprocessing

110 000 numerical experiments are carried out. Approximately 50 % of the total numerical experiments are assigned to two wall materials: gypsum and concrete. The experiments also include different fire conditions with peak HRR ranging from 100 kW to 5100 kW and time to peak HRR ranging from 100 s to 1000 s. In addition, the experiments accounts for different bedroom door opening time, ranging from 1 s to 900 s and window breakage temperature from 100 °C to 200 °C. The histograms of each parameter for the 110 000 numerical experiments are provided in Appendix A. From these 110 000 cases, 25 853 cases reach the potential flashover conditions (i.e., 600 °C). The flashover data being used to develop P-Flashv2 is available in the NIST data repository[3].

In terms of data structure, each experiment consists of seven temperature signals (S^j). Each signal corresponds to a compartment, j, and the temperature signals from the compartment are denoted as $S_i^j = (s_0, s_1, \dots, s_M)$ where s_0 and s_M are the first and the last temperature for the j experiment, respectively. M is the total duration for each of the experiments, and it is 1930 s. The sampling rate for all temperature signals is 1 s.

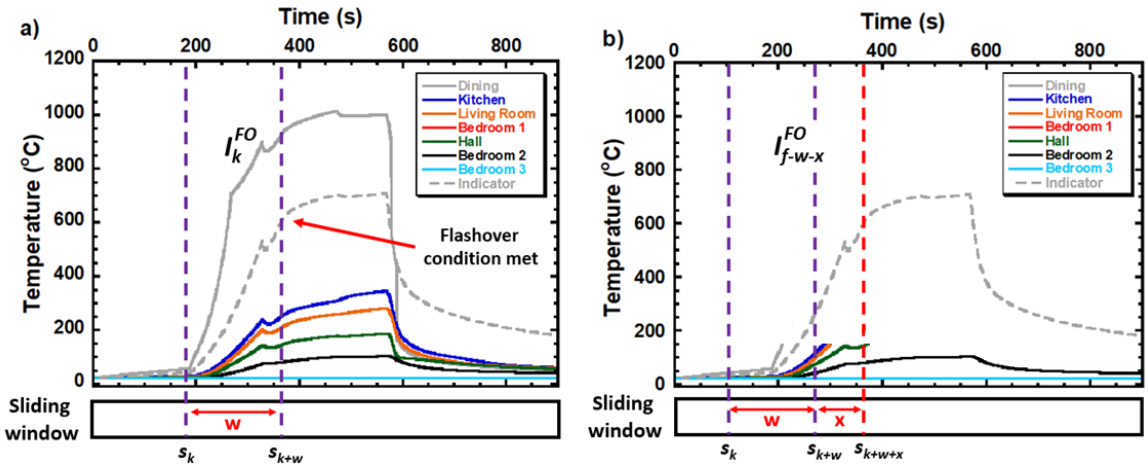


Figure 10: Illustration for a) the use of sliding windows and b) the inclusion of loss of temperature signals and the use of lead time for a fuel-controlled fire case.

3.3.1. Formulation of Instances

Instances need to be constructed for model training/development, and three practical aspects are considered while the instances are formulated. First, the sliding window is applied. The sliding window is used to simulate the physical limits associated with heat detector memory capacity and data transmission. In general, only a certain amount of temperature data can be stored in or transmitted to a computational platform. The schematic of the sliding window is illustrated in Fig. 10a, and it can be seen that an instance from a flashover case is defined as $I_k = \{S_k^1, \dots, S_k^7\}$ where $S_k^j = (s_k, \dots, s_{k+w})$ with k as the first time step of the sliding window and w as the window size.

[3] The temperature dataset can be obtained from the following data repository: https://doi.org/10.18434/mds2-2479. Since the size of the raw data directly from CFAST for the 110,000 cases is more than 2 TB, they will only be provided upon request.

Extracting all the I_k from all fire events, the set of instances $\{[I_k^1, \ldots, I_{k-w}^1], [I_k^2, \ldots, I_{k-w}^2], \ldots, [I_k^{FO}, \ldots, I_{k-w}^{FO}]\}$ are obtained where *FO* is the number of flashover cases. Secondly, the heat detector temperature cut-off is applied. Given a specified threshold mentioned in section 2.1.5, the heat detector failure moment (t_b) for signal S^i is first determined. Then, if time $t \geq t_b$, the temperature, s_t, at time *t* and onward is replaced by a value of 0 °C to represent a loss of temperature signals. An example is provided in Fig. 10b where a temperature cut-off at 150 °C is applied. It should be noted that the dash line for the dining room temperature is only to reference temperature to determine when a flashover condition is met. Thirdly, lead time is considered for the prediction of the potential occurrence of flashover. Due to movement limit (i.e., crawling to avoid excessive heat from ceiling), it will take 10 s for fire fighters to travel approximately 3 m in a fire scene [59]. For that, making predictions ahead of flashover occurrence is crucial in order to allow the fire fighters to get away from the dangerous compartments or find shelter. Figure 10b shows a flashover data instance, I_{f-w-x}^{FO}, when sliding window and lead time are both applied. In the present study, two lead time settings: *x* = 30 s and x = 60 s, are considered.

3.3.2. Labeling of Instances

Each instance is labeled to form our data samples. Since the present study focuses on the prediction of potential flashover occurrence based on temperature data, the input instances are the temperature data and the outputs are the corresponding targets which can either be *Flashover* or *Non-Flashover*. In the current dataset, there is a data imbalance for samples associated with two targets. It can be seen that when *x* = 30 s with a sampling interval of 15 s, there are only 2 Flashover samples in one fire event (i.e., $[I_{f-w-15}^{FO}$ and $I_{f-30}^{FO}]$ where f is the flashover moment for fire event *FO*). Likewise, when *x* = 60 s, the number of Flashover samples is 4 for one fire event (i.e., $[I_{f-15}^{FO}$, I_{f-30}^{FO} , I_{f-45}^{FO}, $I_{f-60}^{FO}]$). However, there are many Non-Flashover samples. In order to overcome the data problem, we first take all the Flashover samples and randomly select the Non-Flashover samples from each fire event. These samples are used to form the final dataset. In this setting with a sampling interval of 15 s, there are a total 103 412 and 206 824 data samples (from 25 853 fire events where the flashover condition is met) for x = 30 s and *x* = 60 s, respectively.

3.3.3. Assigning Training and Testing Subsets

The final dataset is sub-divided into three subsets: training, validation, and testing sets. Specifically, data samples from a set of 2585 fire events are assigned to both the validation (2585) and testing (2585) sets. The remaining 20 683 fire events (25 853 – 2585 - 2585) are given to the training set. The relative portion of the data assignment is identical to both *x* = 30 s and *x* = 60 s. In order to facilitate the training process, mini-batches are used, and the batch size is determined to be 2048 data samples. Using the Adam optimizer [55] with a learning rate of 0.75e-4, the neural network weights and biases from the above equations are updated accordingly. SoftMax cross-entropy [55] is used to measure the loss.

4. Results and Discussion

4.1. Model Performance Against Other RNNs

Table 5 shows the approximate model performance for the prediction of potential occurrence of flashover for P-Flashv2 with attentional BiGRU and eight other RNN architectures with a lead

time of 30 s and 60 s. The data are obtained from the 25 853 flashover cases (see Sec. 3.3) with the sliding window size of 5 min, a flashover threshold of 600 °C, and the heat detector maximum operational temperature limit of 150 °C. The comparison consists of four measures: accuracy, precision, recall, and F1 score. In principle, the accuracy is the ratio of the number of correct predictions to the total number of predictions. The precision is the ratio of the true positive to the sum of the true positive and false positive, and recall is the ratio of the true positive to the sum of the true positive and false negative. If a model has more false positive numbers (i.e., classify flashover for non-flashover cases), the precision scores lower. Likewise, if the recall score is low, this indicates that the model classifies flashover cases as non-flashover. In general, these two measures provide additional details about the model prediction capabilities. The F1 score provides another means to access the model accuracy that gives more weight to false negatives and false positives while not letting large numbers of true negatives influence the score. When data is imbalanced, the F1 score is particularly useful.

Table 5: Model performance of P-Flashv2 against eight different RNN architectures.

x	Model	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
30 s	Basic RNN	81.2	81.0	81.6	81.3
	LSTM	83.6	82.7	85.0	83.8
	GRU	83.7	77.2	95.7	85.5
	LSTM-ATT	82.7	86.7	77.3	81.7
	GRU-ATT	84.9	78.2	***96.9***	86.5
	BiLSTM	84.6	80.7	90.9	85.5
	BiGRU	84.0	78.5	93.7	85.4
	BiLSTM-ATT	87.2	***82.9***	93.7	88.0
	BiGRU-ATT	***87.7***	81.9	***96.9***	***88.8***
60 s	Basic RNN	86.6	***88.8***	83.8	86.2
	LSTM	84.3	78.7	93.9	85.7
	GRU	84.3	81.6	88.6	85
	LSTM-ATT	86.6	***88.8***	83.9	86.2
	GRU-ATT	89.3	84.9	95.6	89.9
	BiLSTM	86.3	82.3	92.4	87.1
	BiGRU	84.1	79.0	92.9	85.4
	BiLSTM-ATT	89.0	85.1	94.5	89.6
	BiGRU-ATT	***89.5***	84.5	***96.0***	***90.0***

For cases with a lead time setting of 30 s, it can be seen from Table 5 that the proposed model using BiGRU with the attention mechanism (BiGRU-ATT) has the highest overall prediction accuracy of 87.7 %. Yet, it should be noted that the model using BiLSTM with the attention mechanism (BiLSTM-ATT) also has the same level of overall prediction accuracy with ~ 87.2 %. BiLSTM-ATT has a slightly higher precision score which indicates that the model would have the least number of false alarms for flashover. For recall score, BiGRU-ATT has ~ 96.9 %, and this result suggests that BiGRU-ATT can effectively recognize the potential occurrence of flashover. Besides the comparison of model performance, it is worth noting that BiGRU-ATT is more

lightweight (i.e., smaller network structure with one less gated mechanism) as compared to BiLSTM-ATT. For that reason, BiGRU-ATT will provide a more numerically efficient model for practical applications.

For cases with a lead time setting of 60 s, the overall prediction accuracy for most of the models including BiGRU-ATT and BiLSTM-ATT is improved. Also, results show that models with the attention mechanism out-perform the stand-alone RNN models. It should be noted that the difference in overall accuracy is due to the difference in the total number of instances from the two settings (lead time of 30s and 60 s). Readers can see Appendix B for the prediction accuracy for each instance from the two settings. However, it is important to highlight that when the lead time becomes much larger, the overall accuracy will begin to drop because the input data are at room temperature. The effect of the attention mechanism will be discussed below.

Fig. 11 illustrates the learned attention weights between heat detector temperature signals of room of fire origin and other compartments. There are two different fire cases. Fig. 11a shows a kitchen fire case and Fig. 11b shows a dining room fire case. In each fire case, there are two bedroom door opening conditions: opened doors and closed doors. The abbreviations K, D, L, H, B1, B2, and B3 are for the kitchen, dining room, living room, hallway, bedroom1, bedroom2, and bedroom3, respectively. It can be seen in both figures that the attention-based model can discover the spatial relation of temperature signals from different compartments. For example, when a fire occurs in the kitchen with all bedroom doors closed, the temperature signals of the dining room, living room, and the hallway are determined as the most effective surrogates (see Fig. 11a). Similarly, when a fire occurs in the dining room with all bedroom doors closed, the most effective surrogate temperatures for flashover predictions are from the kitchen, living room, and the hallway. Physically, this makes sense because these four different compartments (D, K, L, and H) are very close to each other. For the situations when all bedroom doors are all open, the temperature signals from the bedrooms do provide information to the model for flashover prediction, but it can be seen that the learned attention weights from the bedrooms are typically smaller. It should be noted that the difference for the learned attention weights between different bedrooms is driven by the door opening time and the effect of door opening time to temperature importance for predictions is described below.

Fig. 12 shows the learned attention weights between the heat detector temperature signals of the room of fire origin and other compartments for two different fire cases in bedroom 2. The fire setting is similar but the door opening sequence and time are different. For the fire case as shown in Fig. 12a, the door of bedroom 1 and 3 are first opened at ~ 142 s and then the bedroom 2 door is opened at ~ 352 s. For another fire case as shown in Fig. 12b, the bedroom 2 door is first opened at ~ 84 s and the doors for bedroom 1 and 3 are opened later at ~ 189 s. It can be seen in Fig. 12 that learned attention weights for the hallway, living room, dining room, and kitchen are significantly different. When the bedroom 2 door is opened earlier, temperature signals from the kitchen and dining room are shown to be useful for predicting flashover. For cases with all bedroom doors being closed, only temperature signals from bedroom 2 are useful. These observations indicate that, even when no information about the interior door conditions is provided, the attention-based model can determine the relationships between different heat detectors under different door opening conditions and discriminate temperature signals with less importance. Therefore, the model can provide more reliable flashover predictions.

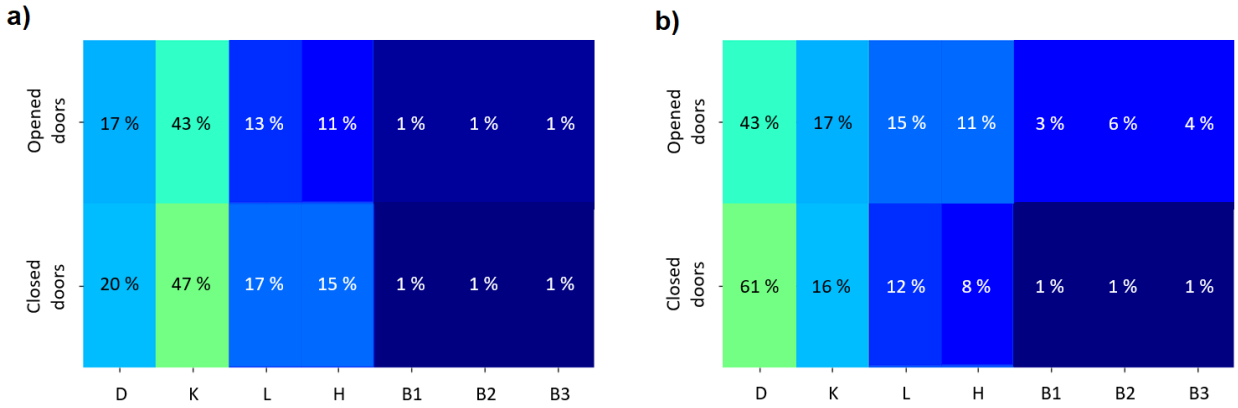


Figure 11. Learned attention for a) kitchen fire and b) dining room fire under different door opening conditions.

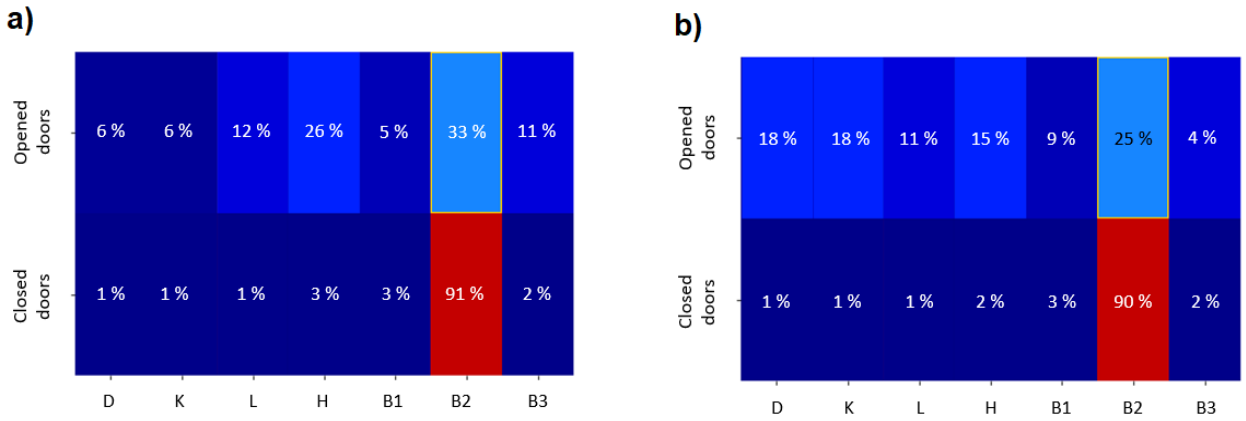


Figure 12. Learned attention for bedroom 2 fire cases with bedroom 2 door being opened at a) ~ 352 s and b) ~ 84 s.

4.2. Model Testing Against Real Fire Data

Given the fact that P-Flashv2 is developed based on synthetic temperature data, it is necessary to examine its performance against real fire scenarios. In this evaluation process, 13 sets of full-scale experiments reported in [36] are utilized. The building structure is identical to that shown in Fig. 1. In these experiments, a single item is first ignited in either the living room, kitchen, or bedroom 1. It should be noted that flashover occurred in most of the experiments, except Exp 6 and 8. Table 6 provides a summary of the primary fire location, ignited item, and ventilation conditions for each fire experiment.

Temperature measurements approximately 2 cm below the ceiling from different compartments are used. Fig. 13 depicts the complete temperature profiles of each compartment for two fire tests: 1) Exp 1: a living room fire experiment with a sofa as the first ignited item and all doors and windows are closed and 2) Exp 13: a bedroom 1 fire experiment with a mattress as the first ignited item, two open doors (bedroom 1 and front door) and, open windows in bedroom 1. The use of temperature signals near the ceiling is to mimic the local temperature that typical heat detectors would measure. Also, although the complete temperature profiles are shown in Fig. 13, only

temperatures below 150 °C (i.e., to account for heat detector maximum temperature operational limit) will be used as model inputs. Based on videos recorded from the thermal image cameras from [36], the time for the transition to flashover is being identified for all fire experiments. The transition to flashover for Exp 1 happens at around 180 s, but flashover does not occur in Exp 6. As shown in Fig. 13, the rapid temperature increase from the room of fire origin can also be observed from the indicator temperature, indicating a fully involved room fire which is a typical fire phenomenon from flashover. The indicator temperatures are measured approximately 0.6 m above the floor surface. It should be noted that P-Flashv2 has never seen any of these experimental data during the training process. From these comparisons, insights regarding data fluctuation, local temperature effects, and the use of CFAST as data generator can also be examined.

Table 6. Fire location, first ignited item, and ventilation conditions for each experiment.

Exp #	Fire Location	Ignited item	Ventilation Settings	Flashover
1	Living Room	Sofa	All Vents Closed	Yes
2	Living Room	Sofa	All Vents Closed	Yes
3	Living Room	Sofa	Front Door Open	Yes
4	Living Room	Sofa	Front Door Open	Yes
5	Living Room	Sofa	Front Door and Bedroom 3 Window Open	Yes
6	Kitchen	Cabinet	All Vents Closed	No
8	Kitchen	Cabinet	All Vents Closed	No
10	Kitchen	Cabinet	Front Door Open	Yes
11	Kitchen	Cabinet	Front Door Open	Yes
7	Bedroom 1	Mattress	All Vents Closed	Yes
9	Bedroom 1	Mattress	All Vents Closed	Yes
12	Bedroom 1	Mattress	Front Door and Bedroom 1 Window Open	Yes
13	Bedroom 1	Mattress	Front Door and Bedroom 1 Window Open	Yes

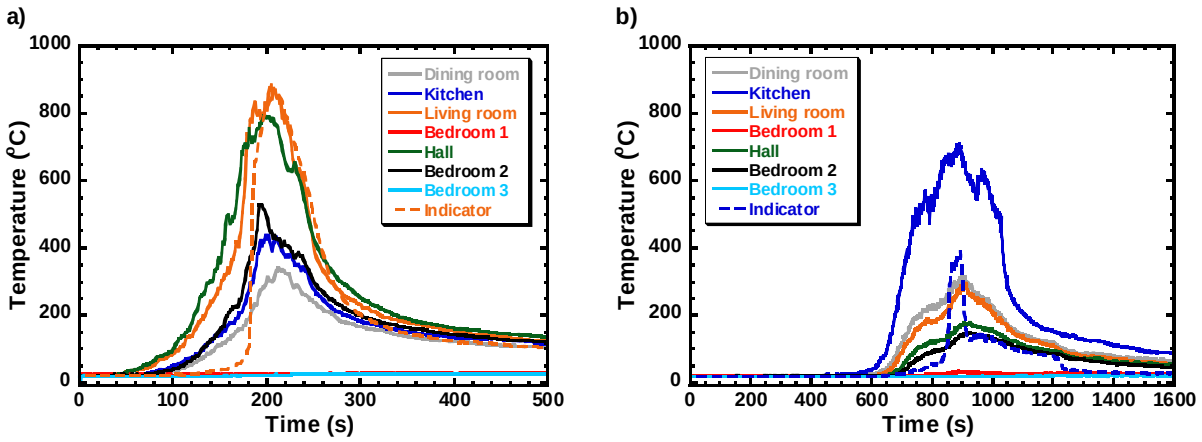


Figure 13. Temperature profiles at different compartments for a) a Living Room fire experiment (Exp 1) and b) a Kitchen fire experiment (Exp 6).

4.2.1. Formulation of Evaluation Instances and Comparison

Due to the nature of the experiments, data associated with flashover and non-flashover conditions are imbalanced. In order to provide a fair comparison, the model performance is assessed based on an equal number of instances. Similar to the data preprocessing described in Sec. 3.1, two non-flashover instances and two flashover instances for the 30 s lead time setting are selected. The longer lead time setting (i.e., 60 s) double the number of instances and so there are a total of eight instances for evaluation from each real fire experiment.

Table 7 shows the approximate overall prediction accuracy for the instances associated with all 13 experiments. P-Flashv2 using BiGRU-ATT provides the best model performance for cases with a lead time setting of 30 s and 60 s, and these results indicate that the model is capable of learning generalized features/temperature information which can be applied to fire scenarios/experiments that the model has never seen. Also, the results suggest that the synthetic temperature data being generated using CFAST are adequate to capture the relevant temperature increase in full-scale real fire experiments for a wide range of fire and vent opening conditions in this single story ranch house.

Table 7. Model performance against real fire data for P-Flashv2 and 6 other models.

	Basic RNN	**LSTM**	**GRU**	**LSTM-ATT**	**GRU-ATT**	**BiLSTM**	**BiGRU**	**BiLSTM-ATT**	***BiGRU-ATT***
Acc. (30s)	74.61	76.92	75.00	76.92	80.76	80.76	79.82	81.76	***82.69***
Acc. (60s)	79.80	80.76	81.73	78.88	83.65	82.69	83.65	84.61	***85.57***

Table 8 and Table 9 show the approximate prediction accuracy at different times for 3 individual experiment series. There are two flashover moments for cases with the lead time setting of 30 s and four flashover moments for cases with the lead time setting of 60 s. Results for non-flashover moments are also given. P-Flashv2 can predict accurately the potential occurrence of flashover ahead of time. For cases with the lead time setting of 30 s, the model performance is excellent and the accuracy for flashover predictions is 100 %. For cases with the lead time setting of 60 s, P-Flashv2 does have one miss prediction in the bedroom 1 experiment case (see Table 9). Yet, the model correctly predicts the flashover moment before 45 s. These results are encouraging because the model provides consecutive correct predictions well before the flashover condition is met.

In terms of non-flashover moments, the overall model performance is good for the living room and bedroom 1 experiment cases. For kitchen fire experiments, the prediction accuracy drops and the significant decrease in performance is primarily due to the fact that the current dataset used from training P-Flashv2 does not cover the kind of kitchen fire scenarios which are included in the full-scale fire experiments. In this study, all fires are located on the floor whereas the ignited wood cabinet in both Exp 6 and Exp 8 from the full-scale fire experiments are located at an elevated location. Fig. 14 provides the temperature profiles for four different kitchen fire experiments. As seen in Fig. 14a and Fig. 14b, the temperature peaks and begins to decay at approximately 900 s for Exp 8 and Exp 10 where all exterior vents are closed. Since heat detectors are assumed to stop functioning at 150 °C and the temperature change in all bedrooms is small, the temperature

decrease is not seen by the model. For that, based on the available temperature information, the model predicts that there could be a potential occurrence of flashover. As seen in Table 10, P-Flashv2 misses 6 out of 8 non-flashover moments and 4 out of 8 non-flashover moments for Exp 6 and Exp 8 (with the lead time setting of 60 s), respectively. Yet, the model performance for the remaining two kitchen fire experiments is very promising. For Exp 10, P-Flashv2 only has one missed prediction and it does not have any missed predictions for Exp 11. Considering temperatures up to 150 °C, it can be seen in Fig. 14 that the temperature growth before flashover for Exp 6 and Exp 8 is very similar to that in Exp 10 and Exp 11. The authors suggest that the current dataset be expanded (i.e., including fire cases with ignited items at elevated locations) in order to enhance model accuracy for cases similar to Exp 6 and Exp 8. The model testing process helped to identify what is required for model improvement.

Table 8. Prediction accuracy of P-Flashv2 using BiGRU-ATT for cases with the lead time setting of 30 s.

	FO - 30 s	**FO - 15 s**	**Non-Flashover**
All	100 %	100 %	70 %
Living room	100 %	100 %	70 %
Kitchen	100 %	100 %	43.75 %
Bedroom 1	100 %	100 %	87.5 %

Table 9. Prediction accuracy of P-Flashv2 using BiGRU-ATT for cases with the lead time setting of 60 s.

	FO - 60 s	**FO - 45 s**	**FO - 30 s**	**FO - 15 s**	**Non-Flashover**
All	91 %	100 %	100 %	100 %	78 %
Living room	100 %	100 %	100 %	100 %	95 %
Kitchen	100 %	100 %	100 %	100 %	68.75 %
Bedroom 1	75 %	100 %	100 %	100 %	95 %

For model efficiency, it should be noted that P-Flashv2 can provide predictions in less than a second. As compared to the current state of the art CFD models such as [18], P-Flashv2 is not only numerically more efficient, but it can also account for realistic fire conditions such as the heat detector temperature limit and various vent opening conditions without the need of prior knowledge. However, P-Flashv2 still has a number of limitations. For example, the model currently relies on the assumption that there is heat detector or heat sensing device at each compartment to have the temperature information. Also, P-Flashv2 can only account for a single building structure with a fixed number of inputs. If there are multiple building structures, the model is now needed to be retrained in order to be applied to another building structure with different number of inputs. Lastly, it is not well understood what specific temperature information the model focuses on making the predications. Therefore, it is necessary to make use of other machine learning techniques, such as feature visualization [61] and/or class activation map [62], to understand the model decision. It is hope that research outcome from this study can contribute a new approach to facilitate real-time data-driven fire fighting and reduce fire fighter deaths and injuries.

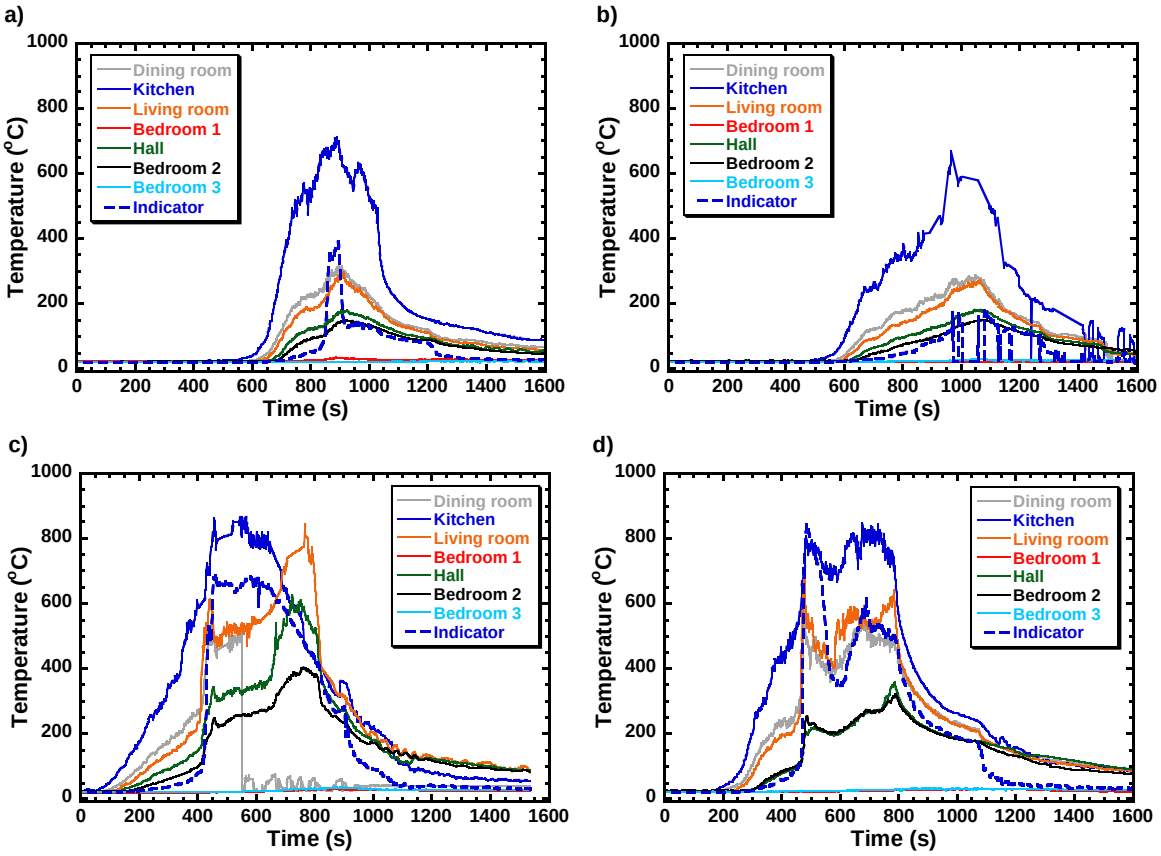


Figure 14. Temperature profiles from different compartments and flashover indicating temperature for a) Exp 6, b) Exp 8, c) Exp 10, and d) Exp 11. Note that there are measurement errors in Exp 6 (indicating temperature) and Exp 8 (dining room temperature).

Table 10. Prediction details for Exp 6, Exp 8, Exp 10 and Exp 11.

	Exp 6	Exp 8	Exp 10		Exp 11	
	Non-FO	Non-FO	Non-FO	FO	Non-FO	FO
Correct Predictions (30 s)	$\frac{2}{4}$	$\frac{2}{4}$	$\frac{1}{2}$	$\frac{2}{2}$	$\frac{2}{2}$	$\frac{2}{2}$
Correct Predictions (60 s)	$\frac{2}{8}$	$\frac{4}{8}$	$\frac{3}{4}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{4}{4}$

Note that the numerator represents the correct prediction and the denominator represents the total number of predictions.

5. Conclusion and Outlook

In this present work, a machine learning based flashover prediction model, P-Flashv2, is developed for a single story ranch-style family home. Using only localized temperature data from heat detectors, P-Flashv2 can predict in advance if there is a potential occurrence of flashover, and is capable of making flashover predictions with a maximum lead time of 60 s. The use of the learning-by-synthesis approach to overcome the lack of training data is discussed. CData is presented to

facilitate synthetic data collection and is used to generate more than 110 000 fire cases with a wide range of fire and vent opening conditions. Machine learning processes and data preprocessing are demonstrated. The model architecture of P-Flashv2 is presented. The proposed model using bidirectional gated recurrent units with attention mechanism (BiGRU-ATT) is benchmarked against 8 different popular recurrent neural network (RNN) architectures. For flashover predictions, results show that with synthetic data, BiGRU-ATT has well-balanced model performance. It has an overall prediction accuracy of ~ 87.7 % (~ 89.5 %) and the recall score of ~ 96.9 % (~ 96 %) for the prediction lead time setting of 30 s (60 s). In addition to model accuracy, P-Flashv2 also has superior model efficiency and as compared to the current state-of-the-art CFD prediction model, P-Flashv2 is at least 300 times faster (< 1 sec vs > 300 sec). P-Flashv2, developed based on synthetic data, is tested against 13 different real fire scenarios. The overall model performance is good, but the current dataset needs to be expanded in order to capture oxygen limited fire scenarios with ignited items at elevated locations. Results from this study demonstrate that P-Flashv2, the machine learning based model, has the potential to facilitate real-time data-driven fire fighting and help reduce fatalities and injuries for fire fighters.

6. Acknowledgements

The authors wish to acknowledge Dr. Daniel Madrzykowski and Dr. Craig Weinschenk from the Underwriters Laboratories Fire Safety Research Institute for providing the valuable full-scale experimental data.

7. References

[1] Babrauskas, V. and Krasny, J.F., 1985. Fire behavior of upholstered furniture (NBS Monograph 173).

[2] Mowrer, F.W. and Williamson, R.B., 1987. Estimating room temperatures from fires along walls and in corners. Fire Technology, 23(2), pp.133-145.

[3] Poon, L.S., 1988. Predicting time of flashover. Fire Safety Science, 3, pp.283-294.

[4] 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.

[5] Babrauskas, V., Peacock, R.D. and Reneke, P.A., 2003. Defining flashover for fire hazard calculations: Part II. Fire Safety Journal, 38(7), pp.613-622.

[6] Yun, K., Bustos, J. and Lu, T., 2018. Predicting rapid fire growth (flashover) using conditional generative adversarial networks. arXiv preprint arXiv:1801.09804.

[7] W ang, Z., Zhang, T., Wu, X. and Huang, X., 2022. Predicting transient building fire based on external smoke images and deep learning. Journal of Building Engineering, 47, p.103823.

[8] Mozaffari, M.H., Li, Y. and Ko, Y., 2023. Real-time detection and forecast of flashovers by the visual room fire features using deep convolutional neural networks. Journal of Building Engineering, 64, p.105674.

[9] Li, Y., Ko, Y. and Lee, W., 2022. RGB image-based hybrid model for automatic prediction of flashover in compartment fires. Fire safety journal, 132, p.103629.

[10] Lin, C.C. and Wang, L.L., 2017. Real-time forecasting of building fire growth and smoke transport via ensemble kalman filter. Fire Technology, 53(3), pp.1101-1121.

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

[12] 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.

[13] Hägglund, B., 1980. Estimating flashover potential in residential rooms. FOA Rapport C, 202369, p.A3.

[14] 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.

[15] Davis, W.D., Cleary, T., Donnelly, M. and Hellerman, S., 2003. Using sensor signals to analyze fires. Fire Technology, 39(4), pp.295-308.

[16] 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.

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

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

[19] Hamins, A.P., Bryner, N.P., Jones, A.W. and Koepke, G.H., 2015. Research roadmap for smart fire fighting (Special Publication (NIST SP)-1191).

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

[21] Jahn, W., Rein, G. and Torero, J.L., 2011. Forecasting fire growth using an inverse zone modelling approach. Fire Safety Journal, 46(3), pp.81-88.

[22] Peacock, R.D., Reneke, P.A. and Forney, G.P., 2017. CFAST–Consolidated Model of Fire Growth and Smoke Transport (Version 7) Volume 2: User's Guide. NIST Technical Note 1889v2.

[23] National Fire Protection Association, 2010. NFPA 72: national fire alarm and signaling code. The Association.

[24] Campbell, R., Evarts, B., and Molis, J.L. 2019. Firefighter Injuries in the United States – 2018, Report, National Fire Protection Association, Quincy, MA.

[25] Fahy, R. F., Petrillo, J. T., and Molis, J. L. 2020. Firefighter Fatalities in the United States - 2019, Report, National Fire Protection Association, Quincy, MA.

[26] Soleymani, M., Pantic, M. and Pun, T., 2011. Multimodal emotion recognition in response to videos. IEEE transactions on affective computing, 3(2), pp.211-223.

[27] Zhao, Z.Q., Zheng, P., Xu, S.T. and Wu, X., 2019. Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning Systems, 30(11), pp.3212-3232.

[28] Venkateswarlu, R., 2003, October. Eye gaze estimation from a single image of one eye. In Proceedings Ninth IEEE International Conference on Computer Vision (pp. 136-143). IEEE.

[29] Geng, X., Zhou, Z.H. and Smith-Miles, K., 2007. Automatic age estimation based on facial aging patterns. IEEE Transactions on pattern analysis and machine intelligence, 29(12), pp.2234-2240.

[30] Asuncion, A. and Newman, D., 2007. UCI machine learning repository.

[31] 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.

[32] 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.

[33] Wu, X., Park, Y., Li, A., Huang, X., Xiao, F. and Usmani, A., 2020. Smart Detection of Fire Source in Tunnel Based on the Numerical Database and Artificial Intelligence. Fire Technology.

[34] McGrattan, K., Hostikka, S., McDermott, R., Floyd, J., Weinschenk, C. and Overholt, K., 2013. Fire dynamics simulator user's guide. NIST Special Publication, 1019(6).

[35] 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.

[36] Madrzykowski, D. and Weinschenk, C. 2019. Impact of fixed ventilation on fire damage patterns in full-scale structures. Underwriters Laboratories Firefighter Safety Research Institute. Columbia, MD.

[37] Mattern, J. L., 2017. It's Official: Americans Are Obsessed with Ranch Homes. https://www.countryliving.com/real-estate/a44039/ranch-home-united-states/. Accessed: 2021-01-020.

[38] Babrauskas, V. and Peacock, R.D. 1992. Heat release rate: the single most important variable in fire hazard. Fire Safety Journal 18(3): 255-272.

[39] Lawson, R., Walton, D. and Twilley, W., 1983. Fire Performance of Furnishings as Measured in the NBS Furniture Calorimeter Part 1. National Bureau of Standards, NBSIR 83-2787. Gaithersburg, MD.

[40] Babrauskas, V., 1979. Full-Scale Burning Behavior of Upholstered Chairs. National Bureau of Standards, TN 1103. Gaithersburg, MD.

[41] 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.

[42] 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.

[43] Babrauskas, V., 2011. Glass breakage in fires. Fire Science and Technology, Inc.

[44] 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. Springer.

[45] 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.

[46] Peacock, R.D., McGrattan, K.B., Forney, G.P. and Reneke, P.A., 2017. CFAST—Consolidated Fire and Smoke Transport (Version 7)—Volume 3: Verification and Validation Guide. NIST Technical Note 1889v3, CFST Version, 7(0).

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

[48] NFPA 72 National Fire Alarm Code 2002 Edition. National Fire Protection Association. Quincy, Massachusetts.

[49] Svozil, D., Kvasnicka, V. and Pospichal, J., 1997. Introduction to multi-layer feed-forward neural networks. Chemometrics and intelligent laboratory systems, 39(1), pp.43-62.

[50] Kalchbrenner, N., Grefenstette, E. and Blunsom, P., 2014. A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188.

[51] Connor, J.T., Martin, R.D. and Atlas, L.E., 1994. Recurrent neural networks and robust time series prediction. IEEE Transactions on Neural Networks, 5(2), pp.240-254.

[52] Minh, D.L., Sadeghi-Niaraki, A., Huy, H.D., Min, K. and Moon, H., 2018. Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network. IEEE Access, 6, pp.55392-55404.

[53] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017, December. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (pp. 6000-6010).

[54] Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural Networks, 4(2), pp.251-257.

[55] 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 symposium on operating systems design and implementation (16) (pp. 265-283).

[56] Chung, J., Gulcehre, C., Cho, K. and Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

[57] Graves, A. and Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 18(5-6), pp.602-610.

[58] Guo, L., Li, N., Jia, F., Lei, Y., and Lin, J. 2017. A Recurrent Neural Network based Health Indicator for Remaining Useful Life Prediction of Bearings. Neurocomputing 240: 98-109.

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

[60] Bottou, L., 2014. From machine learning to machine reasoning. Machine Learning, 94(2), pp.133-149.

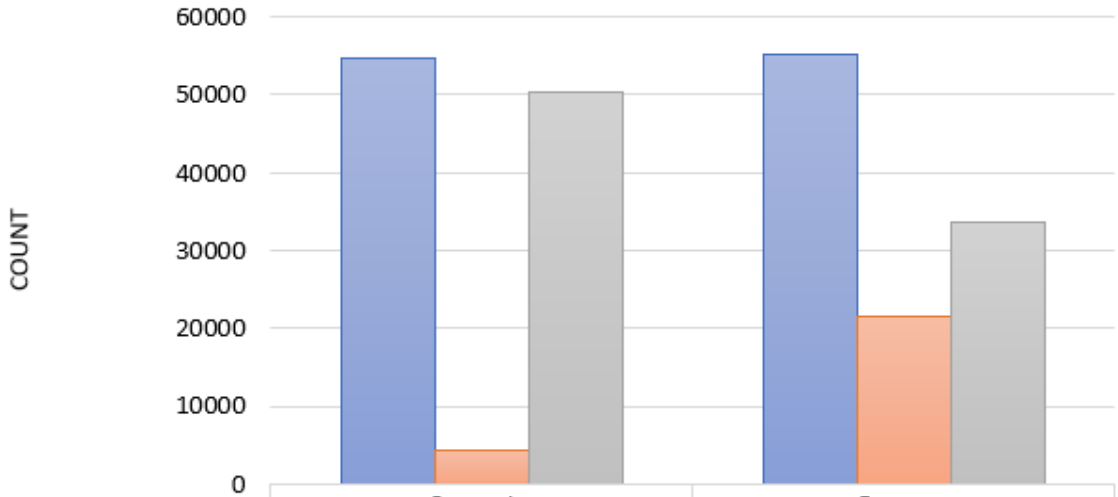
[61] Zeiler, M.D. and Fergus, R., 2014. Visualizing and understanding convolutional networks. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13 (pp. 818-833). Springer International Publishing.

[62] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626).

Appendix A: Data Distribution for Development of P-Flash

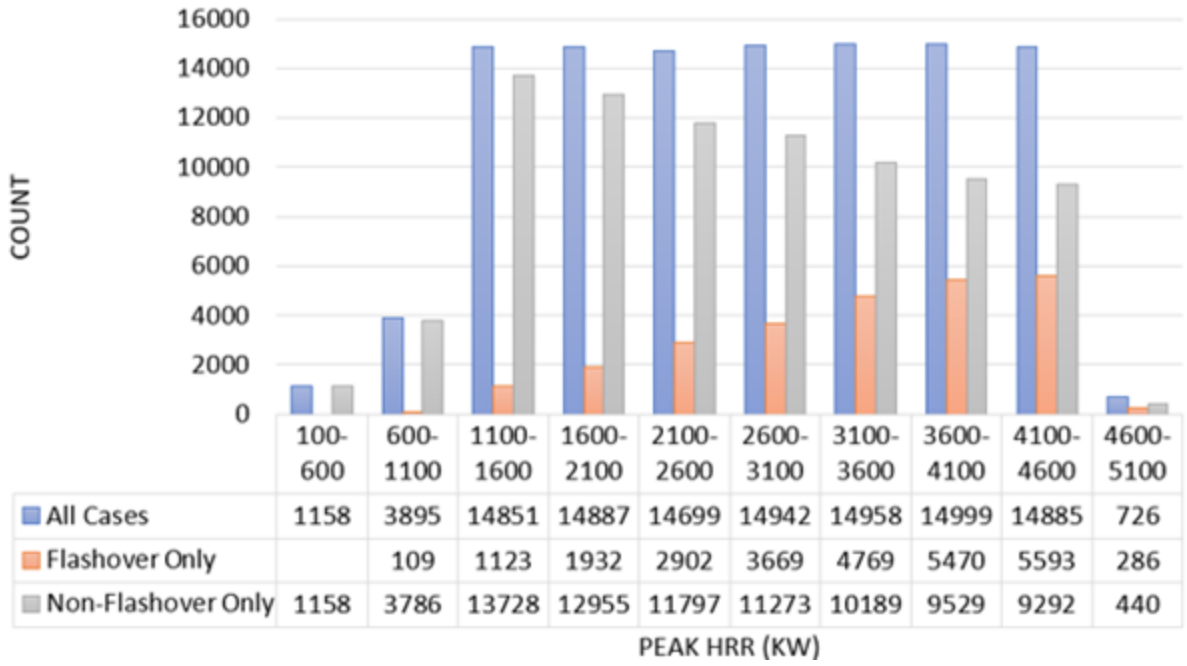
This appendix provides the histograms for eight varying parameters for generating the complete set of data for the development of P-Flash. Namely, they are 1) materials, 2) time to peak HRR, 3) peak HRR, 4) fire locations, 5) fire of origin, opening conditions for 6) exterior doors, 7) interior doors, and 8) exterior windows. There are three sets of colored columns: blue, orange, and gray representing all cases, flashover only cases, and non-flashover only cases, respectively.

In total, 110,000 cases are completed. The size of the raw data is approximately two terabytes. Extracting the cases where the temperature meets the threshold for potential occurrence of flashover (i.e., upper layer gas temperature from the room of fire origin is at least 600 °C), the number of cases is reduced to 25,853 with a reduced data size of roughly five gigabytes. In principle, the histograms/data distributions are provided to present the range of fire and vent opening conditions being included for P-Flash. The reduced data set can be obtained at the data repository: doi:10.18434/mds2-2479, and the raw dataset can only be provided upon request. Note that this dataset can be used to benchmark the performance of more sophisticated machine learning flashover prediction models.



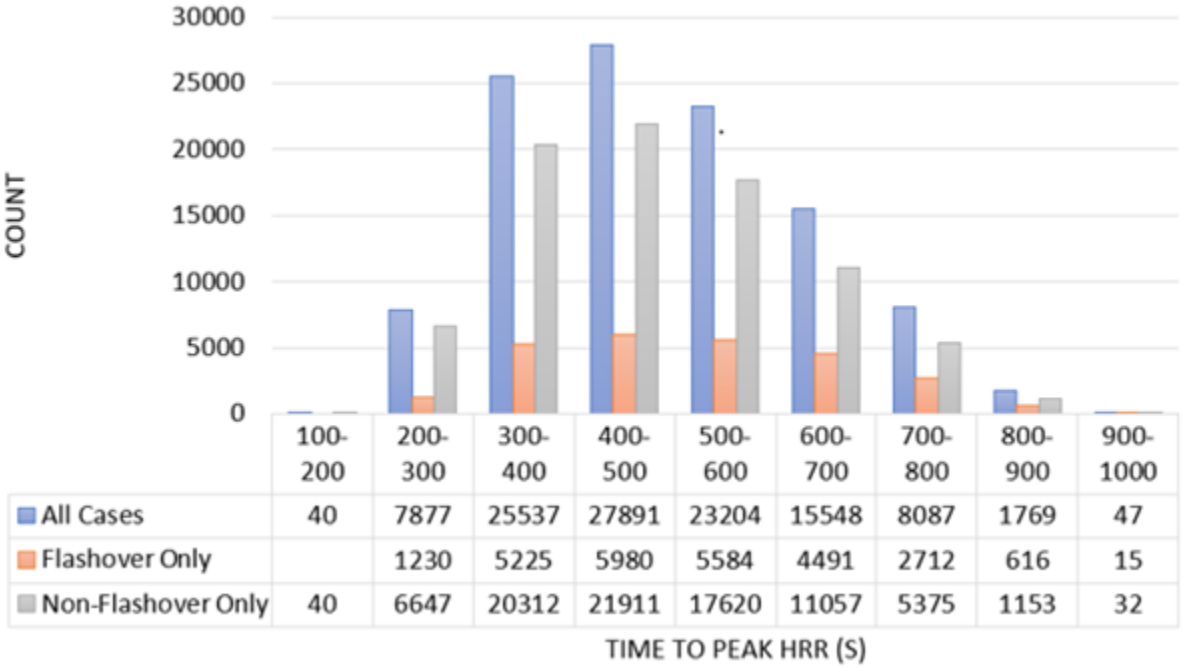
	Concrete	Gypsum
All Cases	54802	55198
Flashover Only	4341	21512
Non-Flashover Only	50461	33686

Figure A1. Histogram of building materials.



	100-600	600-1100	1100-1600	1600-2100	2100-2600	2600-3100	3100-3600	3600-4100	4100-4600	4600-5100
All Cases	1158	3895	14851	14887	14699	14942	14958	14999	14885	726
Flashover Only		109	1123	1932	2902	3669	4769	5470	5593	286
Non-Flashover Only	1158	3786	13728	12955	11797	11273	10189	9529	9292	440

Figure A2. Histogram of peak HRR.



	100-200	200-300	300-400	400-500	500-600	600-700	700-800	800-900	900-1000
All Cases	40	7877	25537	27891	23204	15548	8087	1769	47
Flashover Only		1230	5225	5980	5584	4491	2712	616	15
Non-Flashover Only	40	6647	20312	21911	17620	11057	5375	1153	32

Figure A3. Histogram of time to peak HRR.

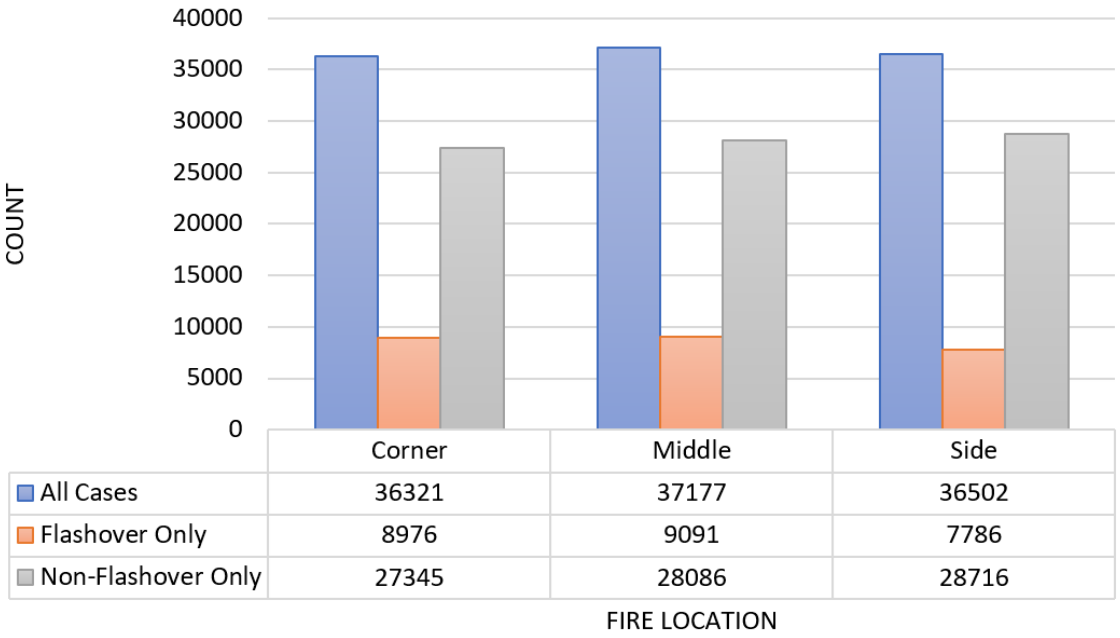


Figure A4. Histogram of fire location.

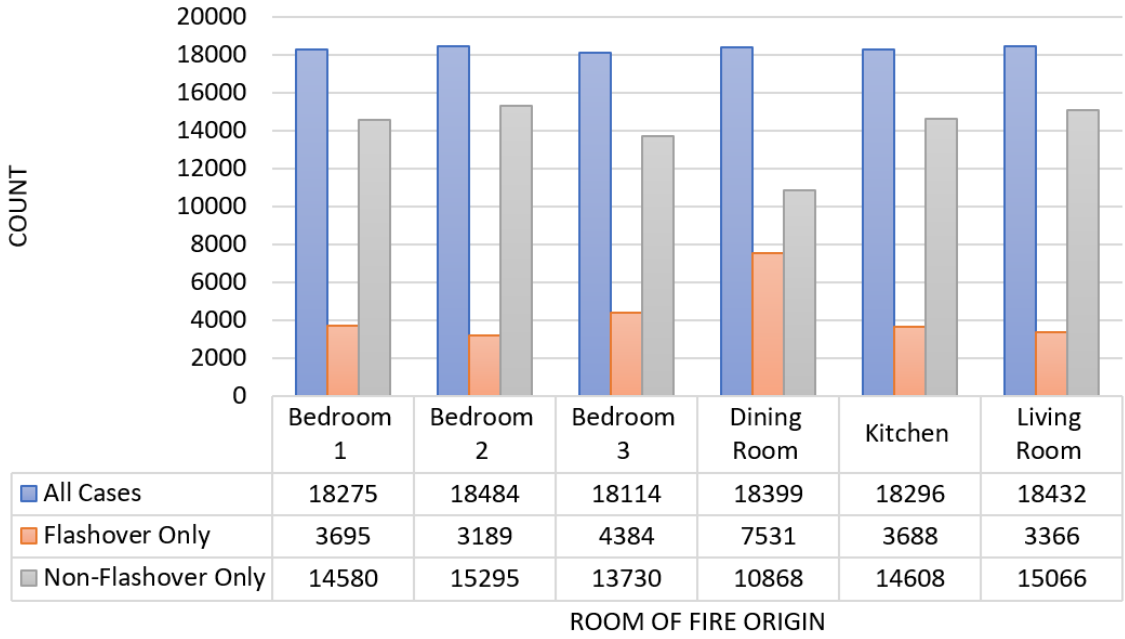


Figure A5. Histogram for room of fire origin.

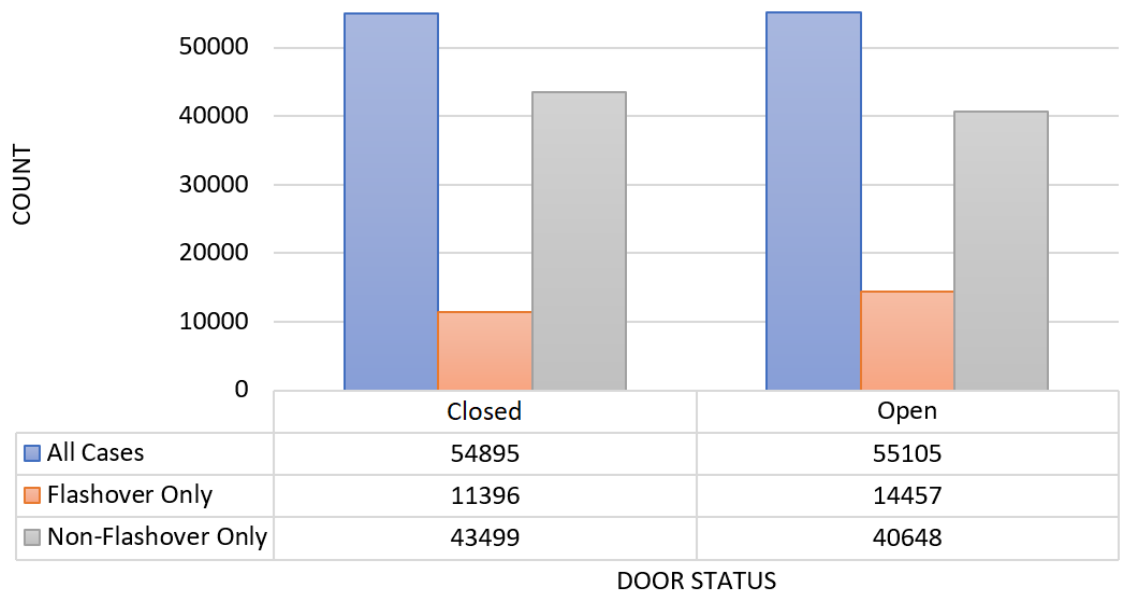


Figure A6. Histogram for door status of front door.

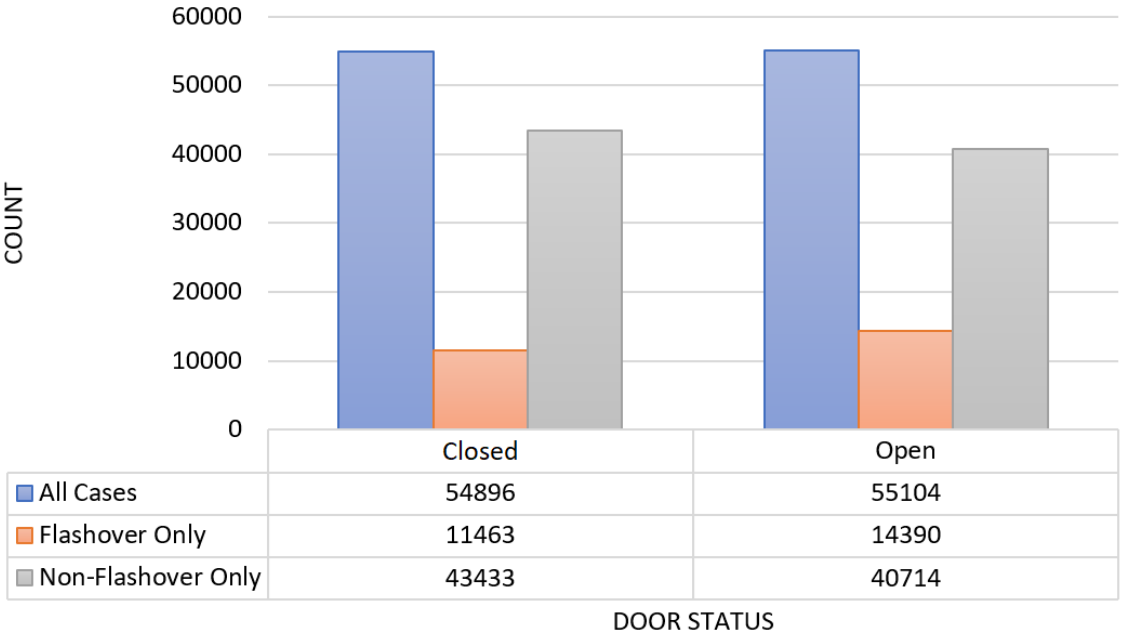
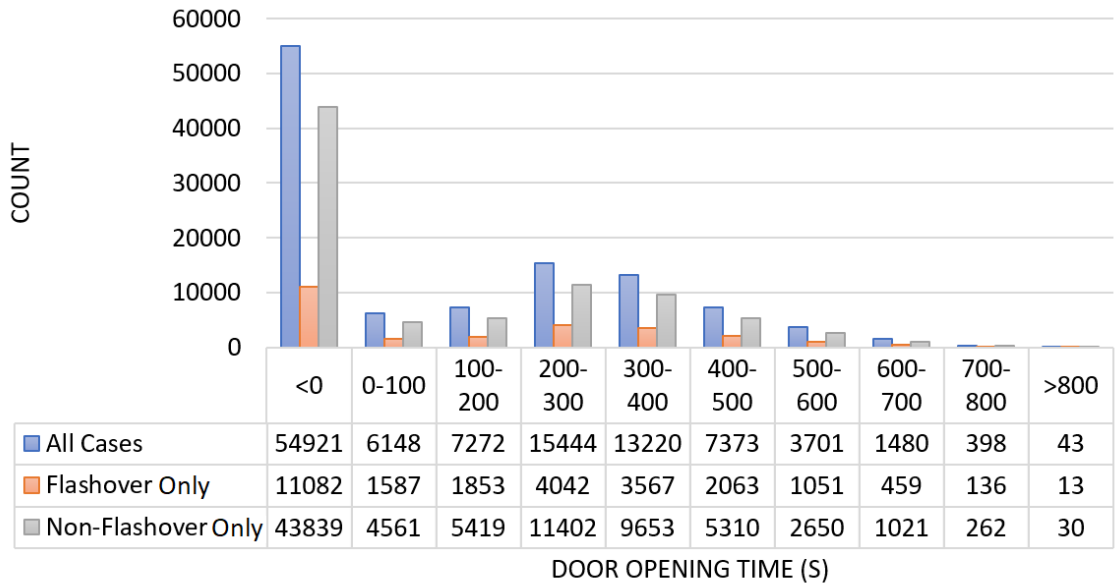
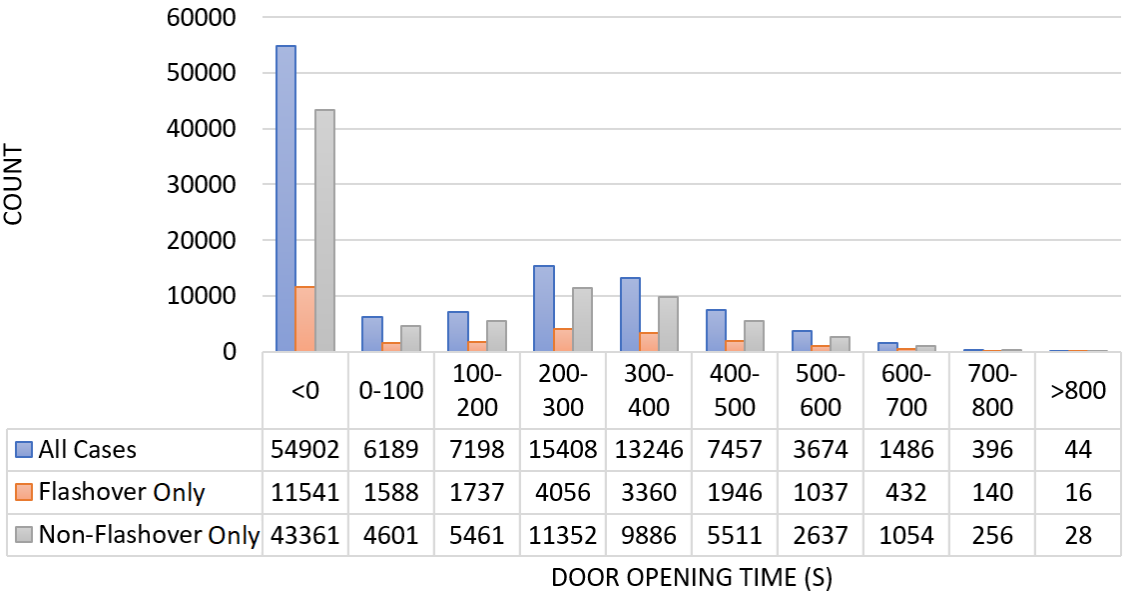


Figure A7. Histogram for door status of kitchen door.



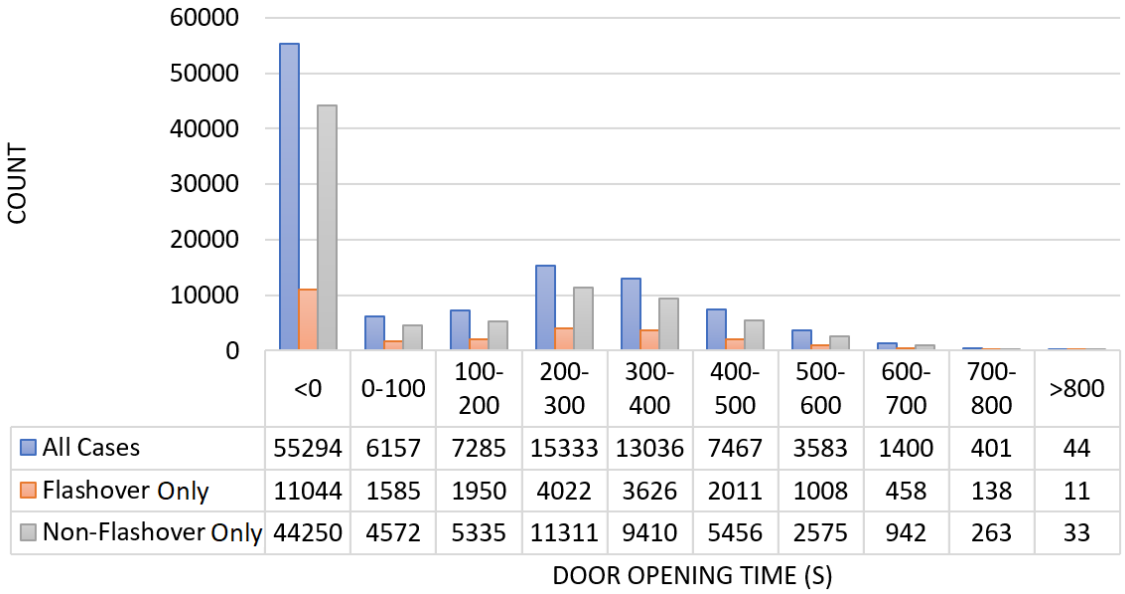
	<0	0-100	100-200	200-300	300-400	400-500	500-600	600-700	700-800	>800
All Cases	54921	6148	7272	15444	13220	7373	3701	1480	398	43
Flashover Only	11082	1587	1853	4042	3567	2063	1051	459	136	13
Non-Flashover Only	43839	4561	5419	11402	9653	5310	2650	1021	262	30

Figure A8. Histogram for door opening time for bedroom 1 door (note that negative time indicates that the door is never opened).



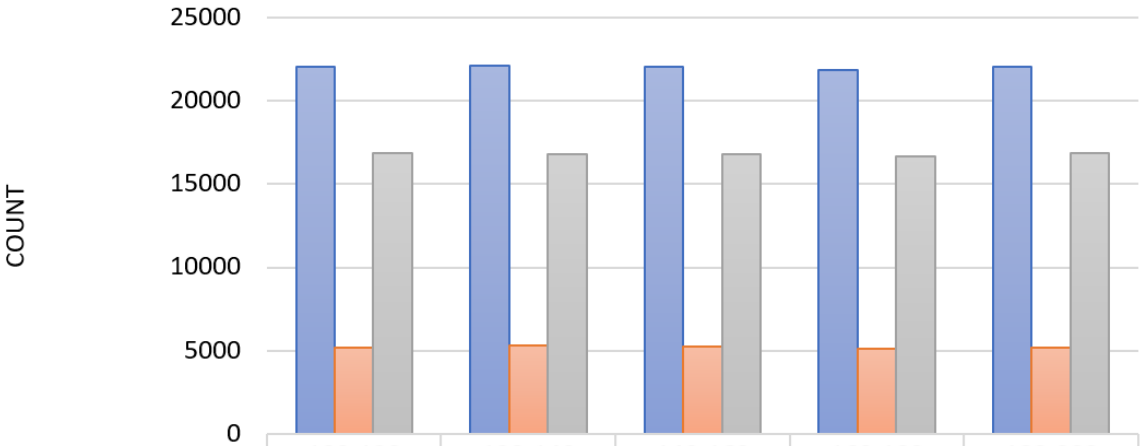
	<0	0-100	100-200	200-300	300-400	400-500	500-600	600-700	700-800	>800
All Cases	54902	6189	7198	15408	13246	7457	3674	1486	396	44
Flashover Only	11541	1588	1737	4056	3360	1946	1037	432	140	16
Non-Flashover Only	43361	4601	5461	11352	9886	5511	2637	1054	256	28

Figure A9. Histogram for door opening time for bedroom 2 door (note that negative time indicates that the door is never opened).



	<0	0-100	100-200	200-300	300-400	400-500	500-600	600-700	700-800	>800
All Cases	55294	6157	7285	15333	13036	7467	3583	1400	401	44
Flashover Only	11044	1585	1950	4022	3626	2011	1008	458	138	11
Non-Flashover Only	44250	4572	5335	11311	9410	5456	2575	942	263	33

Figure A10. Histogram for door opening time for bedroom 2 door (note that negative time indicates that the door is never opened).



	100-120	120-140	140-160	160-180	180-200
All Cases	22020	22094	22028	21816	22042
Flashover Only	5162	5290	5230	5128	5218
Non-Flashover Only	16858	16804	16798	16688	16824

Figure A11. Histogram for window breakage temperature threshold of dining room window.

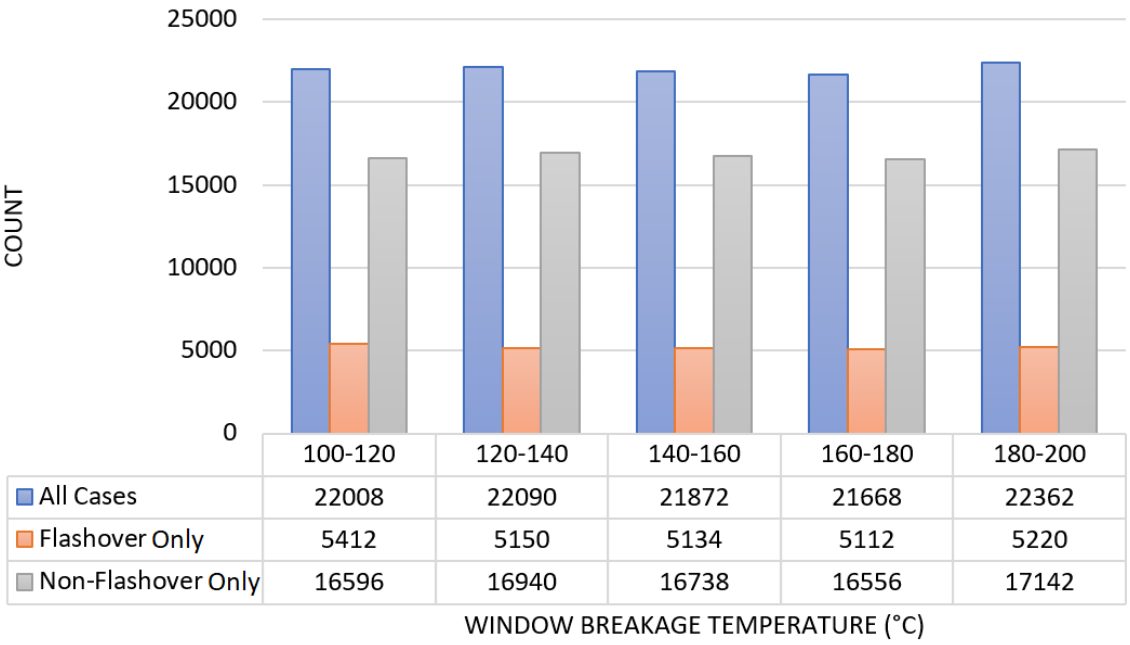


Figure A12. Histogram for window breakage temperature threshold of living room window.

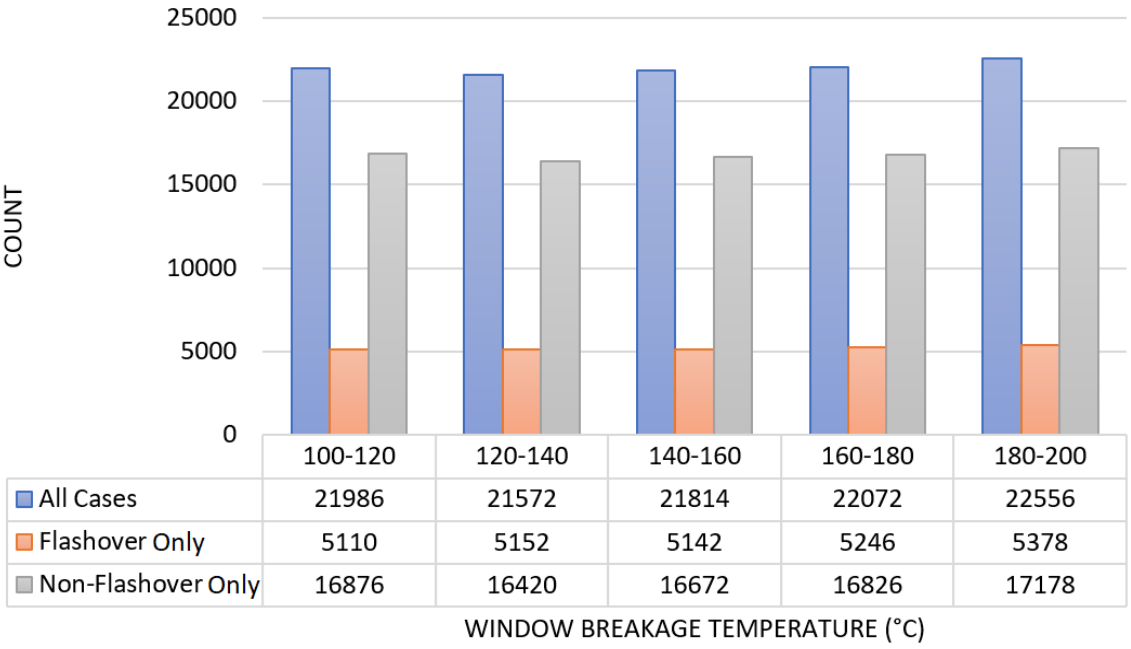


Figure A13. Histogram for window breakage temperature threshold of kitchen window.

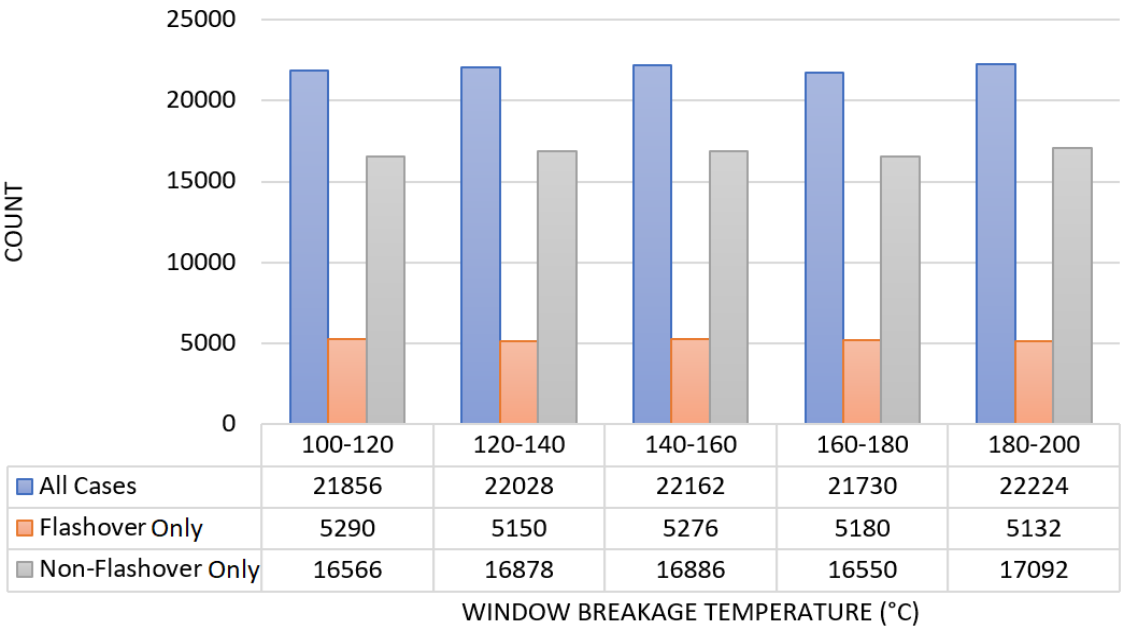


Figure A14. Histogram for window breakage temperature threshold of bedroom 1 front window (E).

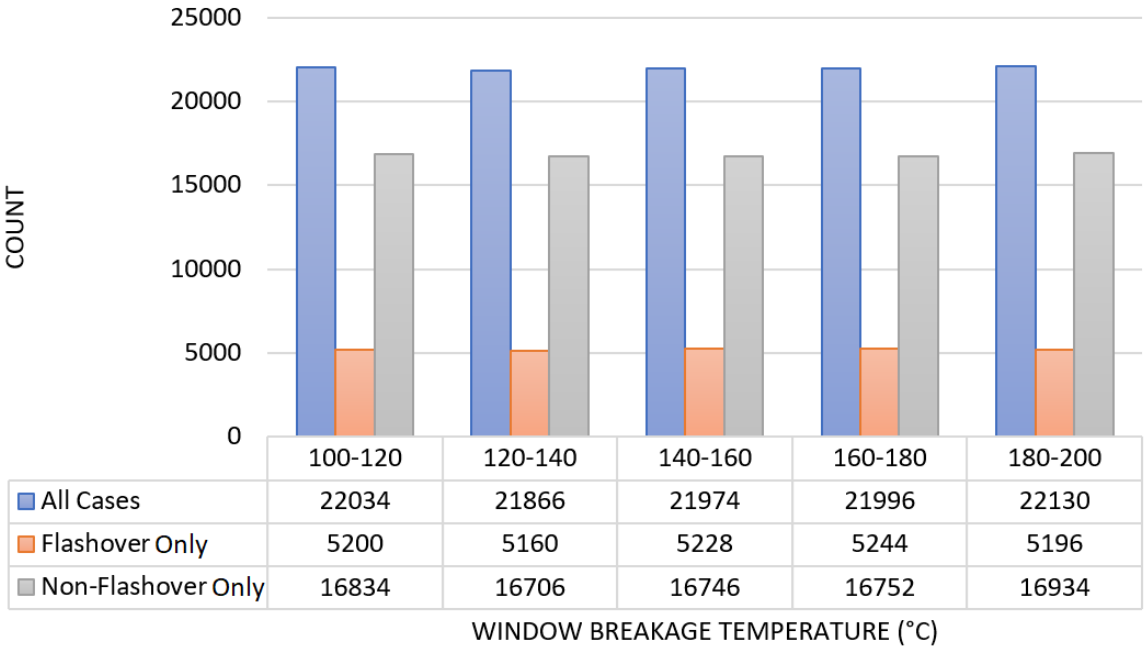


Figure A15. Histogram for window breakage temperature threshold of bedroom 1 side window (D).

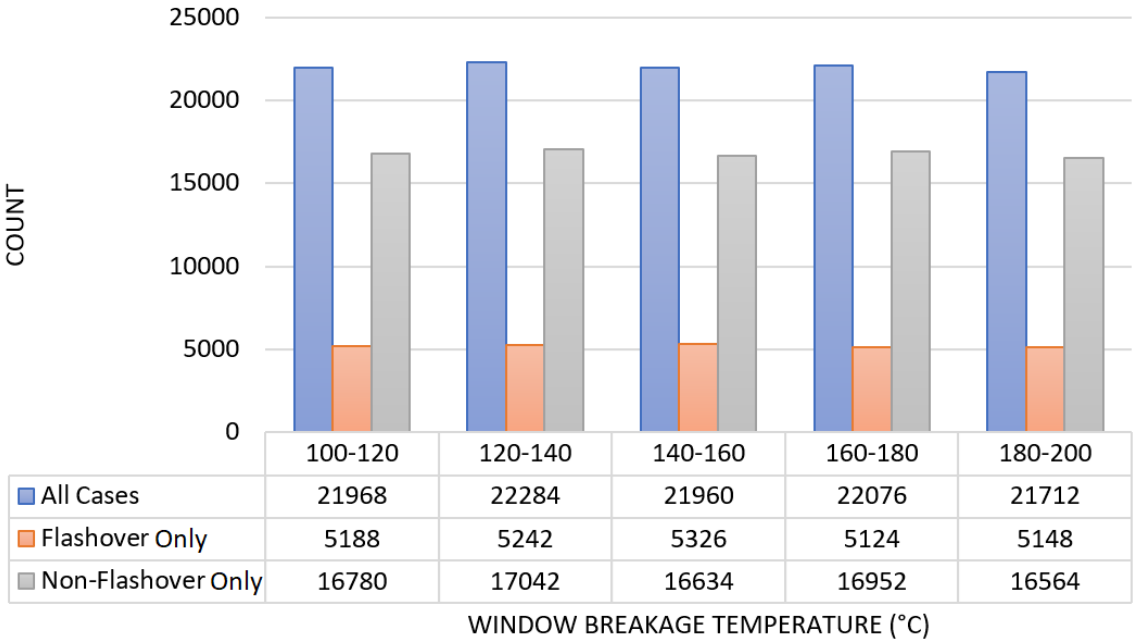


Figure A16. Histogram for window breakage temperature threshold of bedroom 2 window.

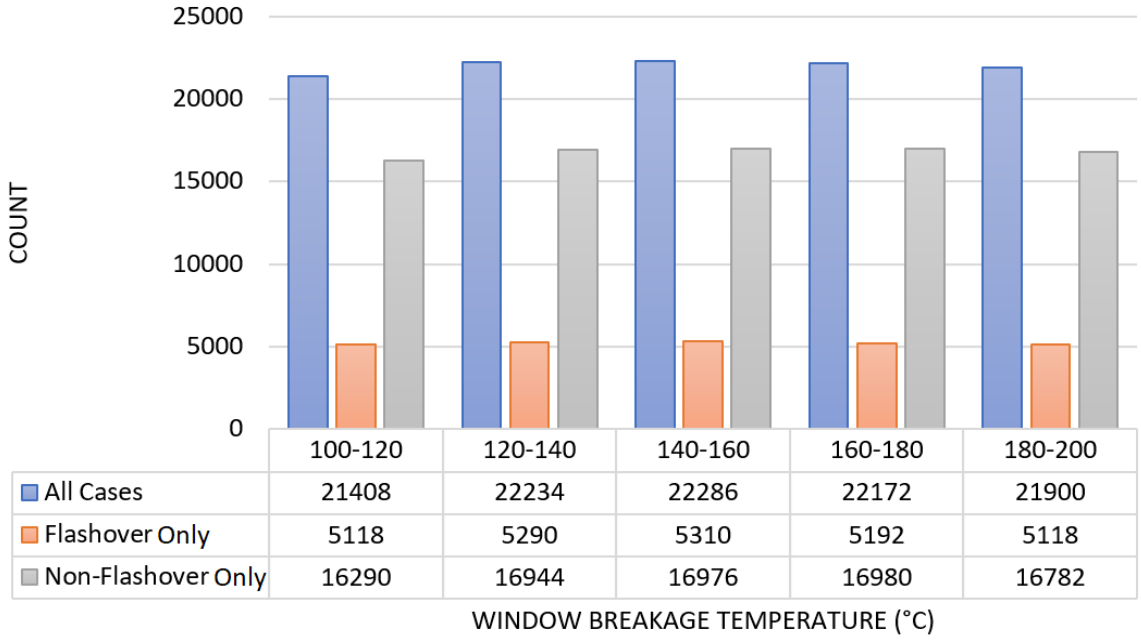


Figure A17. Histogram for window breakage temperature threshold of bedroom 3 window.

Appendix B: P-Flashv2 Prediction Accuracy for Each Instance from Cases with the Lead Time Setting of 30 s and 60 s

For x = 30s cases, the non-flashover instances are 60 s and 45 s and the flashover instances are 30 s and 15 s. It should be noted that the moment that the onset of flashover for the same case regardless of the lead time.

Table B1. Prediction accuracy for each instance from cases with the lead time setting of 30 s.

x = 30 s	**Non-Flashover Instances**		**Flashover Instances**		
	60 s	**45 s**	**30 s**	**15 s**	**Average**
Accuracy	91.9	80.3	84.6	93.8	87.7

As shown from the Table B1, the prediction accuracy tends to drop from 45 s and 30 s instances and this makes sense because the label of these two instances are completely different but the overall temperature data are relatively the same. In generally, the overall accuracy is 87.7 %.

For x = 60 s, the non-flashover instances are (120, 105, 90, 75) s and the flashover instances are (60, 45, 30, 15) s.

Table B2. Prediction accuracy for each instance from cases with the lead time setting of 60 s.

x = 60 s	**Non-Flashover Instances**				**Flashover Instances**				
	120 s	**105 s**	**90 s**	**75 s**	**60 s**	**45 s**	**30 s**	**15 s**	**Average**
Accuracy	89.3	87.4	85.5	79.6	85.0	95.8	96.7	96.9	89.5

As seen from Table B2, there is a similar observation about significant accuracy drop in 75 s and 60 s instances. Also, the model performance for instances away from 75 s and 60 s improves. For that, the average accuracy for all 8 instances is 89.5 %. However, consider only 4 instances similar that in Table B1, the average accuracy (highlighted in red) is about 86.5 % which is comparable to that shown in Table B1 (87.7 %). Therefore, the difference in overall accuracy is due to the number of instances from two different lead time settings.