THE FINAL VERSION OF THE MANUSCRIPT IS PUBLISHED IN THE JOURNAL OF POWER SOURCES: HTTPS://WWW.SCIENCEDIRECT.COM/SCIENCE/ARTICLE/PII/S0378775325005506

Development of a Robust Early-Stage Thermal Runaway Detection Model for Lithium-ion Batteries

Wai Cheong Tam[a,†,*], Jian Chen[b,†], Wei Tang[c], Qi Tong[d], Jun Deng[e], Hongqiang Fang[f], Anthony Putorti Jr.[g,*]

[a]Fire Research Division, Engineering Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland, USA, waicheong.tam@nist.gov

[b]College of Safety Science and Engineering, Xi'an University of Science and Technology, Xi'an, China, chenjian@xust.edu.cn

[c]National Institute for Occupational Safety and Health, 626 Cochrans Mill Road, Pittsburgh, Pennsylvania, 15236, USA, ope0@cdc.gov

[d]Fire Research Division, Engineering Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland, USA, qi.tong@nist.gov

[e]College of Safety Science and Engineering, Xi'an University of Science and Technology, Xi'an, China, dengj518@xust.edu.cn

[f]Fire Research Division, Engineering Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland, USA, hongqiang.fang@nist.gov

[g]Fire Research Division, Engineering Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland, USA, anthony.putorti@nist.gov

[†]Joint first authors, [*]Corresponding authors

Highlights:

- An acoustic-based early-stage thermal runaway detection model is developed.
- Using data augmentation, 2 458 10-s samples are used to develop the detection model.
- The deep-learning detection model never misses any early-stage thermal runaway events.
- The new acoustic-based approach offers near-instantaneous detection of early-stage thermal runaways in near real-life settings.

Abstract:

This paper presents the development of a fast-responding and accurate detection model for early-stage thermal runaway of a lithium-ion battery utilizing acoustics and a deep learning paradigm. A series of single-cell lithium-ion battery tests is conducted. Different state-of-charge conditions and battery orientations are considered. Acoustic data are extracted from video recordings. Using data augmentation, 1 330 acoustic samples of early-stage thermal runaway are obtained. To facilitate the development of a detection model that can be used in real-life settings, 1 128 samples of acoustic data, including various human activities, are also used. Utilizing 10-s acoustic data as the input and a convolutional neural network model structure as the backbone, the detection model has an overall accuracy of 93.9 % with a precision and recall score of 91.6 % and 97.7 %, respectively. Parametric studies are carried out to evaluate the robustness of the proposed model structure and the effectiveness of the data augmentation methods. In addition, the model

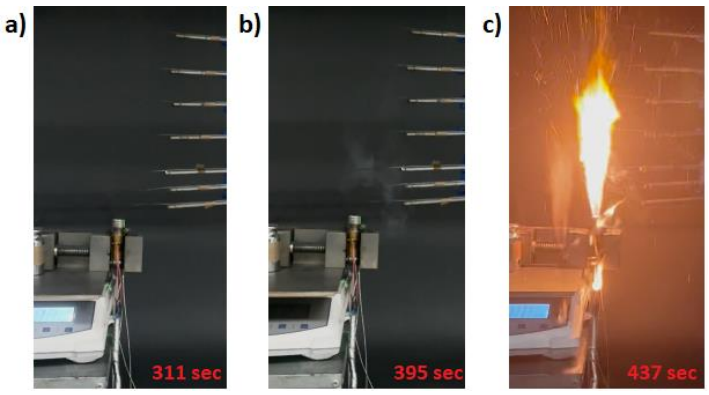
performance against two entire tests is assessed using leave-one test-out cross-validation. It is hoped that the proposed work can help to develop a robust detection device that can provide early warning of thermal runaways and allow users to have extra time to mitigate the potential extreme fire hazards and/or to safely evacuate.

Keywords: thermal runaway prevention; acoustic signals; machine learning; safety valve breakage (SVB); data augmentation; detection

1. Introduction

The use of lithium-ion batteries (LiBs) has been increasingly prevalent in modern society. Today, LiBs are found in nearly every cell phone, laptop, and portable electronic device. The global demand for new LiBs is expected to increase from approximately 700 gigawatt-hours in 2022 to 4 700 gigawatt-hours in 2030 [1]. The benefits of LiBs are profound as they facilitate the development of robust technologies, such as e-bikes, e-scooters, drones, electric vehicles, and energy storage systems. However, the use of LiBs encompasses life-threatening fire hazards that cannot be ignored. In 2023, in New York City, the fire department responded to 267 residential LiB fires which caused 150 injuries and 18 deaths [2]. To illustrate this increasing trend, there were only 10 deaths associated with LiB fires for the combined years of 2021 and 2022. In commercial buildings such as battery warehouses and battery recycling plants, major fires were caused by thermal runaway in LiBs. In 2021, the Superior Battery incident occurred in Morris, Illinois where nearly 4 000 residents were forced to evacuate and millions of taxpayer dollars were spent for environmental clean-up[1]. Based on the injuries, fatalities, and losses, research efforts to address LiB-related fire hazards are needed.

Thermal runaway is a phenomenon in which the lithium-ion cell enters an uncontrollable, self-heating state. Many studies have been conducted to understand various thermal runaway mechanisms, including thermal [3,4], electric [5,6], and mechanical abuse [7,8]. Figs 1 depict three crucial moments of a LiB undergoing a thermal runaway due to thermal abuse: the moment of i) safety valve breakage in Fig. 1a, ii) initial off-gassing in Fig. 1b, and iii) ignition in Fig. 1c. Based on experimental observation, the fire growth of a LiB is extremely fast. Once it is ignited, the fire



Figs. 1. Screenshot for the moment of a) breakage of safety valve (SVB) at around 331 s, b) initial gassing at around 395 s, and c) LiB ignition at around 437 s.

[1] https://www.shawlocal.com/morris-herald-news/news/local/2022/06/28/morris-battery-warehouse-fire-a-look-back-one-year-later/

intensity reaches its peak within about a second, generating a flame jet with peak temperatures of approximately 1100 °C [4]. In the same study, the surface temperature of the LiB was found to be approximately 220 °C. These fire behaviors are dangerous because the high-temperature conditions from the ignited LiB will cause the adjacent LiBs to transition to thermal runaway. Then, the fire will spread rapidly within a minute and become extremely difficult to extinguish, leading to devastating fires such as [2]. For that, a robust approach that can detect the early-stage thermal runaway (i.e., the safety valve breakage as shown in Fig. 1a) is vital because it provides extra time for users/occupants to leave or to mitigate fire hazards which helps to reduce potential injuries, fatalities, and property losses.

Existing fire detection systems cannot detect early-stage thermal runaway of LiBs. In residential and commercial buildings, the commonly used detection technologies are 1) smoke/gas [9,10], 2) heat [11], and 3) video detectors [12,13]. The detection effectiveness of these technologies depends on the fire size and the relative distance between the fire and the detector. As shown in Figs. 1a and 1b, the amount of smoke and hot gases generated from the LiB undergoing a thermal runaway is limited. Unless the detectors are placed directly above the LiB, they will not be triggered to alarm during this early-stage thermal runaway[2]. So, the detection of thermal runaway, even with the ignition of a LiB as shown in Fig. 1c, can be significantly delayed because it takes time for the smoke/hot gases to travel, reach the detector, and arrive in the quantities needed to exceed the triggering threshold [14]. To overcome this detection challenge, sophisticated video detectors with advanced detection models such as [12,13] may be installed. However, there are still two major drawbacks: 1) the video detection system needs a clear line of sight of the targeted LiBs and 2) the systems are extremely expensive. Therefore, the development of a fast-responding approach that can detect the early-stage thermal runaway with visual obstruction is crucial.

The objective of this work is to demonstrate the use of acoustics and machine learning that can facilitate the development of a robust detection model to rapidly detect the LiB thermal runaway at its early stage[2]. For practical engineering applications, acoustic-based detection has been widely used for detecting faulty machine parts [15], gas pipeline leakage [16], and impact damage [17]. In the area of fire research, acoustic-based detection approaches have also been used to identify fire location [18], the heat release rate of fire [19], and the increasing trend of temperature within a compartment [20]. There are two advantages to using acoustic-based detections. Firstly, the signals can be obtained at any moment despite challenging conditions with visual obstruction, limited spacing, and poor lighting. Secondly, because sound travels at the speed of sound, near real-time detection can be achieved [19,20]. However, since the previous studies [15-20] utilize carefully produced acoustic signals (i.e., the signals contain no background noise nor signals derived from any human activities), they will not work in real-life settings. This is because the detection problem becomes more complex if real-life acoustic conditions, such as acoustic signals from human activities and background noises, are considered. It is also worth noting that when the acoustic signals become more complex, traditional rule-based signal processing techniques, such as those described in [20], will not be reliable. Recent literature [21] highlights that the accuracy of a detection model using traditional rule-based techniques degrades significantly as noise levels increase. In order to overcome the technical challenge, the use of a deep learning paradigm is proposed. The primary advantage of using deep learning is that the model can learn the unique acoustic patterns of the safety valve breakage (an indicator of early-stage thermal runaway) and recognize the difference between other impulsive sounds, background noises, and acoustic signals

[2] In this paper, the "early-stage thermal runaway" is denoted as the moment in which the LiB safety valve breaks open.

from human activities. It is believed that the outcome of this work can contribute to a robust detection model that can provide extra time for users/occupants to leave or to mitigate fire hazards, which helps to reduce potential injuries, fatalities, and property losses due to LiB thermal runaway.

The rest of this paper is organized as follows. In Section 2, the experimental setup for data collection, acoustic behavior of the experimental data, gathering acoustic data of other events, and data augmentation for the safety valve breakage acoustic signals are provided. Section 3 describes the development of the early-stage thermal runaway detection model, including the construction of data subsets and model formulation. Section 4 presents the results. Finally, a conclusion and future work are provided in Section 5.

2. Data Collection of Safety Valve Breakage

2.1 Experimental Setup

Figure 2 shows the schematic of the experimental setup which includes an exhaust hood (approximately 2 m x 2 m), a laboratory desk with a heat shield, a battery holder, a lithium-ion battery, a heating element, a glass wall, and a video recorder.

There are two fan speeds for the exhaust hood. A low fan speed has a flow capacity of about 0.2 kg/s and a high fan speed has about 0.45 kg/s. Different fan speeds will yield different levels of background noise. The laboratory desk is positioned at the center of the exhaust hood and it is covered by a stainless steel heat shield to avoid secondary ignition from the battery. The battery holder is located on top of the heat shield. Additional weights are placed on the base of the battery holder to avoid any unnecessary movement. 18650 lithium-ion batteries (LiBs) are used in this study. In terms of properties, the battery cathode is LiNiCoAlO2 and the anode is graphite. The nominal capacity and voltage of the LiBs are 3.2 Ah and 3.7 V, respectively. The LiB is clamp-fastened by the battery holder. Only one type of LiBs is considered in this study. However, the state-of-charge is ranged from 0 % to 100 % with an increment of 25 %. In addition, five battery orientations, which include 0 ° (facing up), 45 °, 90 ° (facing to the side to the page), 135 °, and 180 ° (facing down), are accounted for. A total of 38 experiments are conducted and duplicate

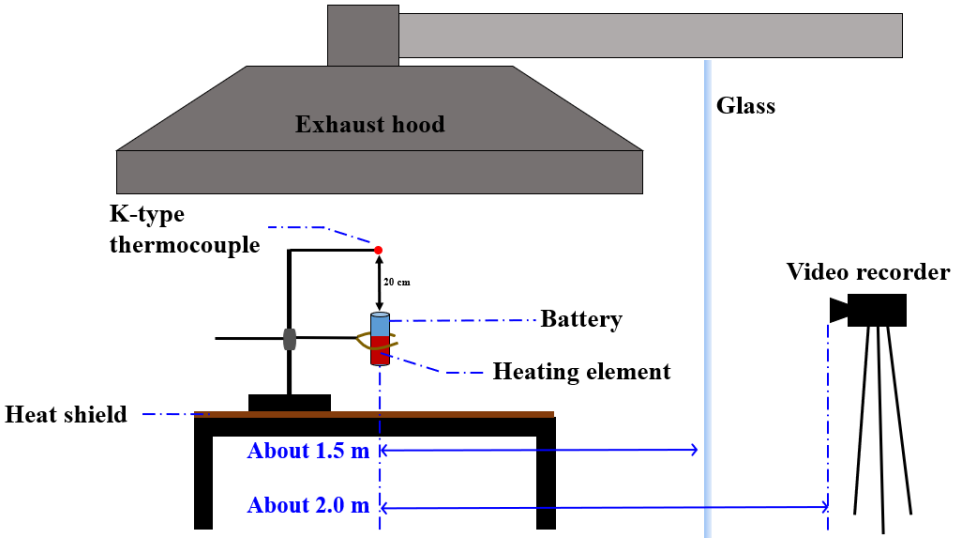
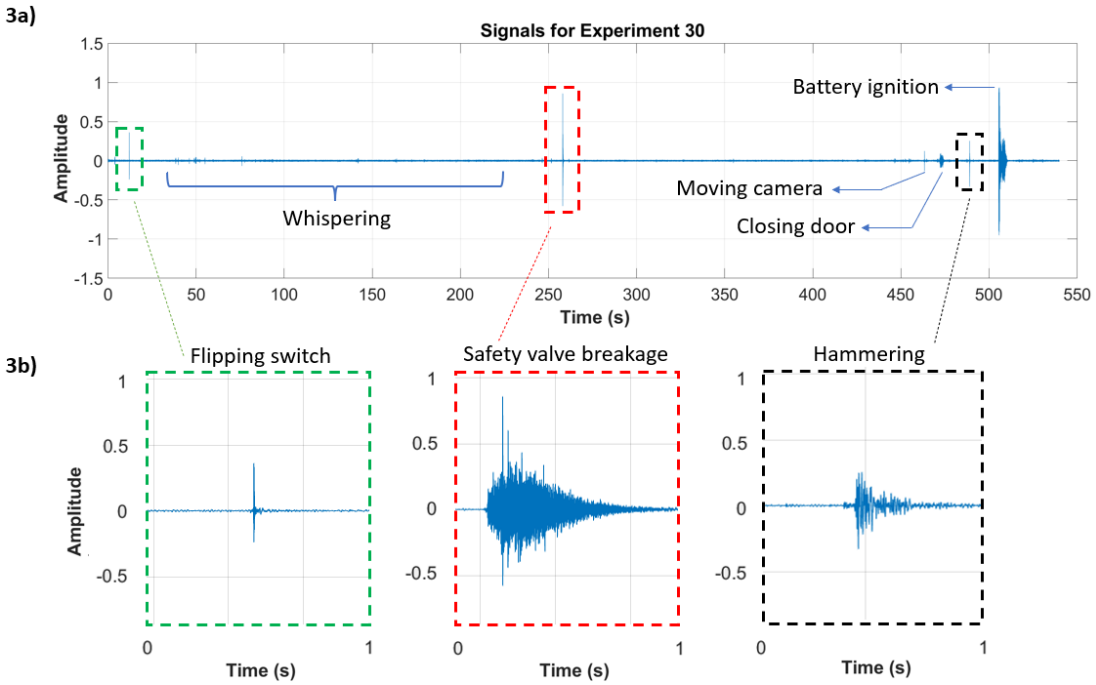


Fig. 2. Schematic of the experimental setup.

experiments are carried out for certain battery configurations (i.e., state-of-charge and orientation). A silicone heating pad powered by nominally DC 12 V and characterized by a 3.8 Ω resistance is used to facilitate the thermal abuse. It has the dimensions of approximately 63 mm x 30 mm. The thickness is approximately 0.25 mm. The heat pad provides a heating temperature range of approximately 175 °C to 200 °C and it is placed onto the lower half of the battery surface. A glass wall approximately 8 mm thick is located approximately 1.5 m away from the battery. It is used as a protective separation. There are gaps on two sides of the glass wall which allow sound to be recorded by the video recorder. FDR-AX700[3] is used to record the experiments and it has a built-in stereo microphone. The duration of each test varies with a mean video record duration of about 473 s (about 8 mins) with a standard deviation of about 126 s. Acoustic signals[4] are extracted from the video recording. The behavior of the acoustic signals will be described in the following subsection.

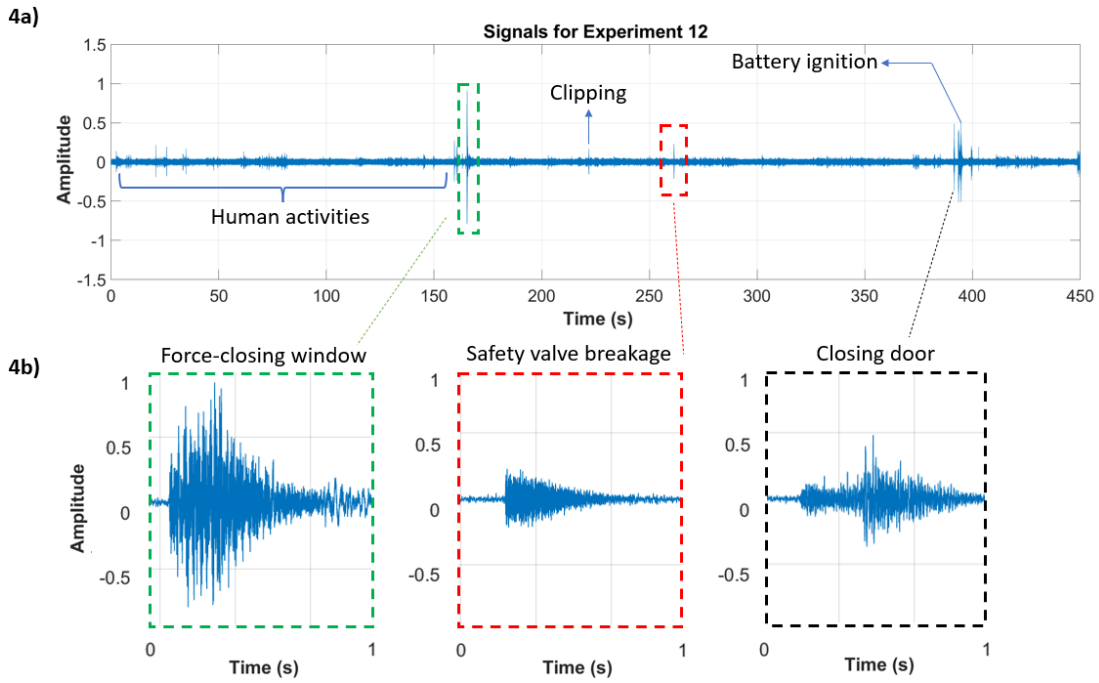
2.2 Experimental Data Behaviors

Acoustic signals of two selected experiments are presented in Figs. 2 and Figs. 3. Fig. 2a shows the acoustic signals for Experiment 30. The safety valve breakage of the LiB happens at around 258 s and the battery ignition takes place at around 504 s. Three important observations are noticed in Fig. 2a. Firstly, background noise is relatively negligible. This is because the fan speed from the exhaust fan is low. Secondly, there are three types of sounds with relatively low amplitude due to various human activities and they are a) whispering, b) adjusting the camera, and c) door closing. Thirdly, there are also four types of acoustic signals with relatively large amplitude and they



Figs. 3. Acoustic data of a) Experiment 30 and b) three selected events.

[3] Certain commercial products or company names are identified here to describe our study adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products or names identified are necessarily the best available for the purpose.
[4] Both channels of the stereo recording are identical. Therefore, channel 1 of the audio recording is used in this study.



Figs. 4. Acoustic data of a) Experiment 12 and b) three selected events.

include i) flipping switch, ii) safety valve breakage (SVB), iii) hammering, and iv) battery ignition. Fig. 2b shows the zoom-in plots of the flipping switch, SVB, and hammering signals. It can be seen from the figure that the SVB signals have unique acoustic shapes and characteristics (i.e., a longer oscillation duration and a larger peak-to-peak amplitude as compared to signals from flipping switch and hammering). It should be noted that each experiment has one SVB event and there are 38 different SVB events observed in all the experiments.

Figure 4a shows the acoustic signals for Experiment 12. The safety valve breakage (SVB) and battery ignition happen at around 261 s and 394 s, respectively. As compared to Experiment 30, there are three major differences. Firstly, the background noise is larger due to an increase in the fan speed from the exhaust hood. Secondly, there is a wide range of human activities which include loud discussion, laughing, walking, preparation for the next experiment, and grinding metal. In addition, Fig. 4a also shows other acoustic signals, such as clipping objects and battery ignition. Finally and most importantly, the SVB acoustic signals are significantly different in Experiment 12. Fig. 4b shows the zoom-in plots for force-closing a window, SVB, and door closing. It can be seen that the acoustic signals of SVB have much lower peak amplitude. Also, the signals begin with maximum oscillations and then decay relatively linearly. In contrast, the SVB signals from Fig. 3b first increase and then decrease exponentially after reaching their peak with a peak amplitude of nearly 0.75. The differences in the signals are likely due to different LiB state-of-charge and its orientation. It is also worth noting that the overall characteristics of acoustic signals for force-closing a window from Fig. 4a are very similar to the SVB shown in Fig. 3b. Since the SVB signals can vary significantly due to different state-of-change and its relative orientation and the fact that the signals from other events can be similar to SVB, the development of a robust early-stage thermal runaway detection model that can be used in real-life settings is challenging.

2.3 Gathering Other Event Acoustic Data

Additional acoustic signals are required to teach the deep-learning detection model to distinguish events other than SVB. To facilitate the learning, three publicly available acoustic datasets are considered and they are TUT [22], AE [23], and ESC [24]. Together, these datasets have approximately 11 900 distinct samples covering a wide range of activities and events. In total, they are about 11 700 mins in length and more than 13 GB in size.

Manual extraction is conducted to obtain a selective dataset. Since the detection model is intended to be used in indoor residential-like settings, only acoustic data of related events are utilized. These events include a) data subsets of Home, Office, and Library from [22] and b) data subsets of pet, speech, non-speech, interior, and exterior noise sounds from [23] and [24]. Table 1 shows the summary of the events from TUT, AE, and ESC datasets. Together, 1 128 data samples for other events (non-SVB samples) are gathered. These data have a sampling rate of 44 100 Hz with a resolution of 24 bit. The duration of these data ranges from approximately 4 s to more than 60 s and the data size is about 5 GB. It should be noted that a more robust deep learning detection model can be developed when the acoustic data from other events are considered. This is because since the detection model has seen these events, the model will be less likely to misclassify these events as SVB. Section 4 will provide supportive results to demonstrate the effectiveness of accounting for these additional data.

Table 1. A summary of events used in TUT [22], AE [23], and ESC [24] datasets.

Dataset	**Event Names**					
TUT	Home		Office		Library	
AE	Guitar	Violin	Cat	Dog	Bird	Child
	Speech			Scream		
	Squeak	Knock	Applause	Rustle	Mouse click	Hammering
ESC	Sneezing	Clapping	Coughing	Laughing	Snoring	Footstep
	Baby crying			Pouring water		
	Toilet flush	Can opening	Wood creaks	Clock alarm	Car horn	Hand saw

2.4 Data Augmentation

Data augmentation is carried out to address the data imbalance problem. As mentioned in Section 2.3, acoustic data from other events are vital to train a robust model. However, there are 1 128 samples for the acoustic data from other events. In contrast, there are only 38 samples for SVB. Training the model with such a biased dataset will be problematic because 1) the model does not have enough SVB sample to learn from and 2) the model is being "forced" to learn the data pattern from other events. In order to resolve this problem, the easiest solution is to reduce the number of other-event samples. Although this will help and the model will be able to recognize SVB events, it will not be able to distinguish non-SVB/other events, such as those from human activities. Therefore, more SVB samples are needed.

Two data augmentation methods are adapted from [25] and they are utilized to generate synthetic SVB samples. Data augmentation is not a new concept. It has been used in many acoustic detection studies. For example, it is shown to effectively boost the model performance in [26]. Guidelines

from [25] are also taken into account such that the label of the synthetic/augmented signals is maintained. The two augmentation methods are described below:

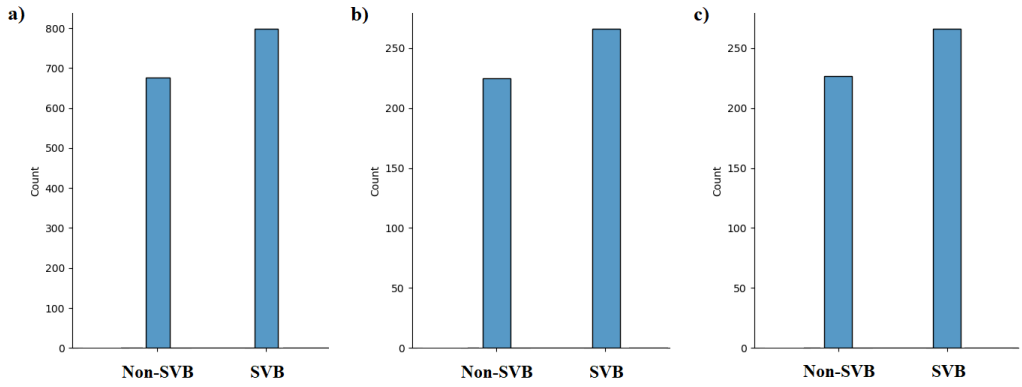
1. Time Stretching (TS): To slow down or speed up the acoustic sample. Each sample is time-stretched by five different factors: [0.81, 0.93, 1.00, 1.07, and 1.19].
2. Pitch Shifting (PS): To raise or lower the pitch of the acoustic sample. Each sample is pitch-shifted by seven different values (in semitones): [−3.5, −2.5, −1, 0, 1, 2.5, and 3.5].

It should be noted that the original acoustic signals are preserved if TS and PS are equal to 1 and 0, respectively. By applying five TS factors and seven PS values to the 38 SVB samples, an additional 1 292 acoustic signals are obtained. It is also worth mentioning that the augmented data can simulate various SVB events for different types of LiBs with different capacities and state-of-charges. In total, there are 1 330 samples for SVB and 1 128 samples for other others. The entire dataset is now more balanced. A parametric study is done to show the effectiveness of data augmentation in Section 4.

3. Model Development for the LiB Early-Stage Thermal Runaway Detection

3.1 Construction of Data Subsets

Four data preprocessing processes are carried out to construct the data subsets that can be used for model training. Firstly, a tumbling window is utilized to extract 10-s samples from the dataset. The use of a tumbling window (i.e., 0 s to 10 s, 11 s to 20 s, etc.) enables the acoustic data of interest to be located anywhere within the 10-s sample. Within the 10-s samples, the acoustic data of interest might also be chopped off and it is believed that the consideration of these data characteristics facilitates the development of a more robust model that can be used in real-life settings. Secondly, padding is used to obtain consistent samples in length. Typical background noises from the TUT dataset are used to pad the samples that have fewer than 10-s of data. Thirdly, z-score normalization is applied to the entire dataset. Finally, a ratio of 60/20/20 is utilized to construct the training, validation, and testing subsets. Fig. 5 shows the distribution of the SVB and other-event/non-SVB samples. It can be seen that the number of SVB and non-SVB samples is relatively even. These data subsets are used to train, validate, and test the deep-learning detection model.



Figs. 5. Data distribution for a) training, b) validation, and c) testing subsets of the entire dataset that contains 2458 samples.

3.2 Detection Model Structure

Motivated by [27,28], a convolutional neural network (CNN) is used as the backbone of the model structure. The model uses a single-channel 10-s acoustic sample as the input. The model output is the corresponding class of the input sample (i.e., non-SVB or SVB).

Figure 6 shows the one-channel convolutional neural network for the detection model. There are four convolutional layers (conv) and each layer consists of a different number of convolutional filters. A stride of 4 is used in the first convolutional layer and a stride of 1 is used for the rest of the convolutional layers. Table 2 describes the configuration details for each layer. In general, the use of convolutional filters offers temporal locality to capture the dynamics of the signals [27]. For the early layers, the filter describes basic shape features, such as a peak, a valley, or a certain degree of slope. In the later layers, where the filter has a larger receptive field, it is more likely to contain shape semantics, such as a large magnitude peak or an impulse with multiple peaks.

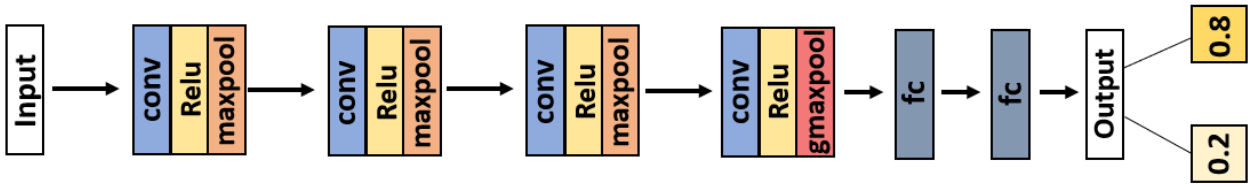


Fig. 6. Overall structure of the detection model.

Two regularization techniques are used to facilitate the training. The first technique is the one-dimensional maximum pooling [29] and there are three benefits: a) to retain the most significant features that can be crucial to determining non-SVB and SVB samples, b) to neglect small distortions such as those due to noise, and c) to reduce the number of elements in the sample going to the next layer. As shown in Table 2, a maximum pooling (maxpool) of 4 is used after the first convolutional layer and a maximum pooling of 2 is used after the second and the third convolutional layers. The second technique is global maximum pooling (gmaxpool) [29]. In contrast to 1-D maximum pooling, the global maximum pooling takes the maximum value from the feature map across all positions. It means that only the most prominent feature is preserved. The advantage is that the exact location of the important feature within the acoustic sample does not matter and so a more robust CNN can be obtained. The exact locations of maxpool and gmaxpool are shown in Fig. 6.

Layer	Type	Output Size	Kernel/ Pooling Size	Stride		Layer	Type	Output Size	Kernel/ Pooling Size	Stride
1	conv	(‡, 110231, 32)	80	4		**4**	conv	(‡, 6883, 32)	3	1
	maxpool	(‡, 27557, 32)	4	–		**5**	gmaxpool	(‡, 32)	–	–
2	conv	(‡, 27555, 32)	3	1		**6**	fc	(‡, 32)	–	–
	maxpool	(‡, 13777, 32)	2	–		**7**	fc	(‡, 16)	–	–
3	conv	(‡, 13775, 32)	3	1		**8**	softmax	(‡, 2)	–	–

‡ This symbol indicates the dimension of the input size.

	maxpool	(⁑, 6887, 32)	2	–						

Table 2. Configuration details for each layer of the detection model.

As shown in Fig. 6, there are two fully connected (fc) layers and they are used to combine the higher lever features to form a classification. The last layer is SoftMax [29] which provides the predicted probability of the two classes (non-SVB and SVB). Fig. 7 presents the loss and accuracy plot for the validation set. The loss function is taken to be the binary cross-entropy. An Adam optimizer [29] with an initial learning rate of 1e-4 was used to optimize the model. Early-stopping [29] with a patience of 50 was used to avoid overfitting. With this configuration, if the loss from the validation subset does not improve for 50 consecutive epochs, the training stops. As seen in Fig. 7, early-stopping happens in Epoch 42 and the model is saved for testing.

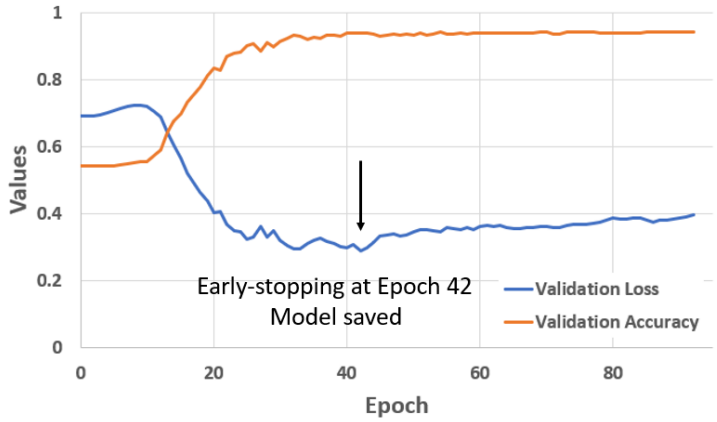


Fig. 7. Validation loss and accuracy plot.

4. Results

4.1 Overall Model Performance

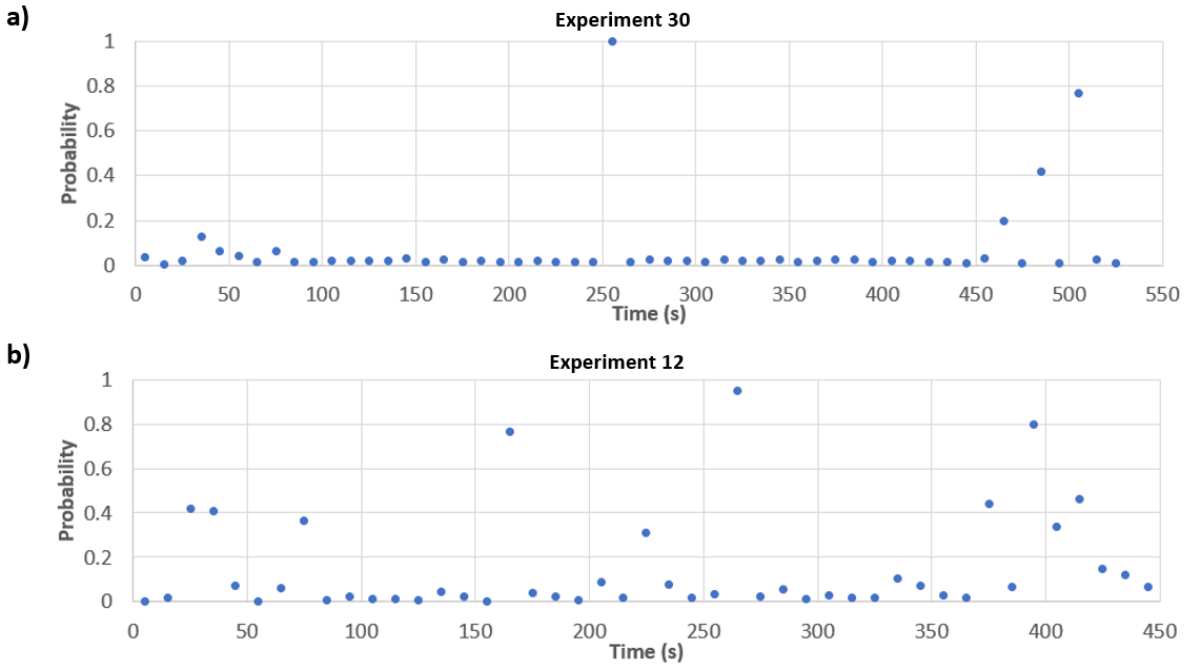
Table 3 presents the performance of the proposed detection model. Three cases are considered. Model assessment is made utilizing the testing subset and the model performance is evaluated based on accuracy, precision, and recall [30]. In the first case, a total of 76 samples are considered. Utilizing the split ratio of 60/20/20 mentioned in Section 3.1, there are 15 testing samples in this case (7 samples are non-SVB and 8 samples are SVB). As shown in the table, the performance of the model (denoted as M14 76) on this dataset is perfect. However, when a wide variety of non-SVB data is included in the dataset, the model performance (denoted as M14 wo DA) drops dramatically. Although accuracy is about 83 %, the scores for precision and recall are zero. This means that the M14 wo DA cannot recognize any of the SVB events. Similar to that mentioned in Section 2.3, since more than 95 % of the data are related to non-SVB events, the model cannot effectively learn to detect SVB. In the last case, when data augmentation is used and the data distribution (see Figs. 5) is more balanced as compared to that used in training M14 wo DA, the performance of the model (denoted as M14 w DA) is significantly better. Although the model performance of M14 w DA is worse than that of M14 76 for the testing subset, the results in the below subsection will highlight the drawback of training the model with only a small set of data.

Table 3. Performance comparison with three model configurations.

	M14 76	M14 wo DA	M14 w DA
Accuracy	100 %	82.7 %	93.9 %
Precision	100 %	0 %	91.6 %
Recall	100 %	0 %	97.7 %
# of Train, Val., Test samples	46, 15, 15	700, 233, 233	1475, 492, 492

4.2 Model Assessment Using Leave-One-Experiment-Out Cross-Validation

Figures 8 present model predictions for each input sample of the entire experiment for Experiment 30 and Experiment 12. The corresponding acoustic signals can be retrieved in Fig. 3a and Fig. 4a, respectively. The inputs are 10-s samples of the experiment from the beginning to the end. The outputs are probabilities of SVB as a function of time for each sample. If the prediction probability is larger than a threshold, the prediction is SVB. Otherwise, the prediction is non-SVB. In this study, the threshold is taken to be 0.5 which is typically being used for binary classification tasks. Leave-one-experiment-out cross-validation (LOEO CV) is used. This means that when the model is tested against Experiment 30, the acoustic data and the augmented data from Experiment 30 will be excluded from the training and validation subsets. The LOEO CV is applied to all 38 experiments. This strategy helps test the model against all possible real-life events that happened during the experiment and identify what the model needs to be improved.

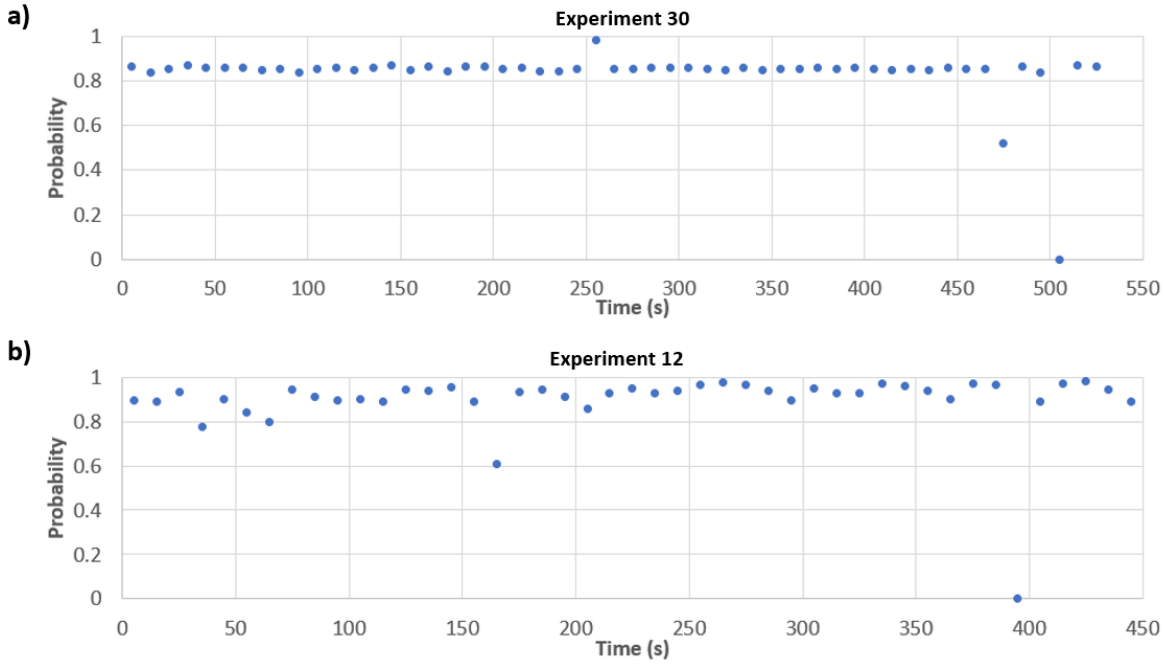


Figs. 8. Predictions of M14 w DA for a) Experiment 30 and b) Experiment 12 with data augmentation.

As shown in Fig. 8a, the model successfully detects the SVB event happened at around 260 s. The prediction can be obtained as soon as the model receives the required 10-s sample. For that, near real-time is achievable and the SVB event can be detected within a couple of seconds. In the same figure, it can be seen that events, such as flipping switches, whispering, adjusting the camera, and door closing, do not affect the model. However, a false positive is observed at around 505 s. This

false positive is corresponding to a battery ignition event and the acoustic signals from any of the battery ignition events are not included in the current dataset. One important takeaway from this testing exercise is that if a practical and robust early-stage thermal runaway detection model needed to be developed, the inclusion of other-event data is crucial.

Figure 8b shows the predictions as a function of time for Experiment 12. The model does correctly detect the SVB event at around 260 s. However, referring to Fig. 4a, it can be seen that this particular experiment contains many non-SVB events. Although the model is capable of distinguishing a wide range of human activities, such as loud discussion and walking and other impulsive sound events, it misclassifies the force-closing window and the combination of door closing and battery ignition events. In addition, the output probabilities for several events at around 30 s, 75 s, 225 s, 375 s, and 410 s are relatively high (i.e., at about 0.4) which indicates that the model sees that some patterns of the sample are highly similar to SVB events. Indeed, additional improvement is needed to help the model distinguish these events and there are two possible strategies. The first strategy is to increase the number of samples, including both non-SVB and SVB events. However, the potential cost can be substantial. Motivated by the drawback of the first strategy, a cheaper or cost-effective approach (i.e., the second strategy) is to make use of higher-level inputs. Rather than using raw acoustic signals, more informative signals, such as Fast Fourier Transform (FFT) [31] and Mel-Frequency Cepstral Coefficients (MFCC) [32], can be used. Both FFT and MFCC provide acoustic signal characteristics in the frequency domain. The second strategy is likely to help the model learn more indicative patterns of SVB without the need to carry out additional experiments. This research effort is currently underway.



Figs. 9. Predictions of M14 wo DA for a) Experiment 30 and b) Experiment 12 without data augmentation.

In order to assess the robustness of the inclusion of more non-SVB data and the use of data augmentation, M14 76 (a developed model using only 76 samples) is tested using LOEO CV for Experiment 30 and Experiment 12. Figs. 9 show the predictions as a function of time. It can be

seen that the model performance is poor. It cannot distinguish non-SVB and SVB events. This kind of model cannot be used in real-life applications. It is suggested that a model interpretability approach, such as [33] or [34], is needed to understand what information the model is using to make predictions. Also, based on this assessment, it is hoped that the use of leave-one-experiment-out cross-validation for the entire dataset of an experiment is vital to fully understanding the deep-learning detection model.

5. Conclusions and Future Work

The development of the deep-learning early-stage thermal runaway detection model using acoustic signals is presented. A total number of 38 thermal abuse tests on a single-cell lithium-ion battery is carried out. From the data study, safety valve breakage (SVB) is identified as the indication of early-stage thermal runaway. Using a tumbling window, 10-s acoustic samples corresponding to SVB are obtained. A convolutional neural network model structure is utilized. Regularization methods, such as maximum pooling, global maximum pooling, and early-stopping, are used to facilitate model training. In order to develop a robust detection model that can be used in real-life settings, three processes are proposed:

1) Including acoustic data from a wide range of human activities, background noises, and pets.
2) Applying data augmentation methods, such as time stretching and pitch shifting, to obtain additional synthetic acoustic samples for SVB.
3) Using leave-one-experiment-out cross-validation (LOEO CV) to fully assess the model performance in near real-life settings.

Based on the results from the parametric studies, there are three major takeaways:

1) It is crucial to have a model that sees sufficient data. With a relatively small dataset (38 non-SVB samples and 38 SVB), although the accuracy of the model (see M14 76) is perfect, such a model with limited knowledge does not work in real-life settings (see Figs. 9). M14 76 cannot distinguish non-SVB and SVB events.
2) The use of data augmentation resolves the data imbalance problem and helps develop a more robust detection model. Having a relatively well-balanced dataset (1 128 non-SVB samples and 1 330 SVB samples, although the model (see M14 w DA) has only an accuracy of 94 % (which is 6 % less than M14 76 as shown in Section 4.1), M14 w DA outperforms M14 76 when the LOEO CV is used to assess the model performance against two selected experiments that have near real-life settings (i.e., human activities and other impulsive sound events).
3) More acoustic data for non-SVB events is needed. In order to enhance the model performance in real-life applications, the future dataset should contain additional acoustic data, such as battery ignition, door closing, and a combination of multiple events.

As for future work, it is believed that the use of higher-level inputs, such as that obtained from Fast Fourier Transform and Mel-Frequency Cepstral Coefficients, will offer a cost-effective strategy for training and will help the model to learn more indicative patterns to accurately distinguish non-SVB and SVB events. This research effort is ongoing and it will be reported in a future publication. The results and observations from this study can help the development of a

robust detection device that can provide early warning of thermal runaways and allow users to have extra time to mitigate the potential extreme fire hazards and/or to safely evacuate such that losses, injuries, and deaths from fires caused by lithium-ion battery thermal runaway can be significantly reduced.

Acknowledgments

The authors from National Institute of Standards and Technology would like to thank the Engineering Laboratory Exploratory Project Program for supporting this research effort.

References

[1] Usai, L., Lamb, J.J., Hertwich, E., Burheim, O.S. and Strømman, A.H., 2022. Analysis of the Li-ion battery industry in light of the global transition to electric passenger light duty vehicles until 2050. *Environmental Research: Infrastructure and Sustainability*, *2*(1), p.011002.

[2] Verzoni, A., 2022. Full Throttle. National Fire Protection Association Journal on Emerging Issues. National Fire Protection Association. Quincy, Massachusetts.

[3] Liu, X., Stoliarov, S.I., Denlinger, M., Masias, A. and Snyder, K., 2015. Comprehensive calorimetry of the thermally-induced failure of a lithium ion battery. *Journal of Power Sources*, *280*, pp.516-525.

[4] Tang, W., Tam, W.C., Yuan, L., Dubaniewicz, T., Thomas, R. and Soles, J., 2020. Estimation of the critical external heat leading to the failure of lithium-ion batteries. *Applied thermal engineering*, *179*, p.115665.

[5] Ohsaki, T., Kishi, T., Kuboki, T., Takami, N., Shimura, N., Sato, Y., Sekino, M. and Satoh, A., 2005. Overcharge reaction of lithium-ion batteries. *Journal of power sources*, *146*(1-2), pp.97-100.

[6] Yuan, Q., Zhao, F., Wang, W., Zhao, Y., Liang, Z. and Yan, D., 2015. Overcharge failure investigation of lithium-ion batteries. *Electrochimica Acta*, *178*, pp.682-688.

[7] Feng, X., Sun, J., Ouyang, M., Wang, F., He, X., Lu, L. and Peng, H., 2015. Characterization of penetration induced thermal runaway propagation process within a large format lithium ion battery module. *Journal of Power Sources*, *275*, pp.261-273.

[8] Kim, C.S., Yoo, J.S., Jeong, K.M., Kim, K. and Yi, C.W., 2015. Investigation on internal short circuits of lithium polymer batteries with a ceramic-coated separator during nail penetration. *Journal of Power Sources*, *289*, pp.41-49.

[9] Chen, S.J., Hovde, D.C., Peterson, K.A. and Marshall, A.W., 2007. Fire detection using smoke and gas sensors. *Fire Safety Journal*, *42*(8), pp.507-515.

[10] Tam, W.C., Fu, E.Y., Mensch, A., Hamins, A., You, C., Ngai, G. and va Leong, H., 2021. Prevention of cooktop ignition using detection and multi-step machine learning algorithms. *Fire safety journal*, *120*, p.103043.

[11] Maksimović, M., Vujović, V., Perišić, B. and Milošević, V., 2015. Developing a fuzzy logic based system for monitoring and early detection of residential fire based on thermistor sensors. *Computer Science and Information Systems*, *12*(1), pp.63-89.

[12] Celik, T. and Demirel, H., 2009. Fire detection in video sequences using a generic color model. *Fire Safety Journal*, *44*(2), pp.147-158.

[13] Qiu, X., Xi, T., Sun, D., Zhang, E., Li, C., Peng, Y., Wei, J. and Wang, G., 2018. Fire detection algorithm combined with image processing and flame emission spectroscopy. *Fire Technology*, *54*, pp.1249-1263.

[14] Mensch, A.E., Hamins, A., Tam, W.C., Lu, Z.J., Markell, K., You, C. and Kupferschmid, M., 2021. Sensors and machine learning models to prevent cooktop ignition and ignore normal cooking. *Fire Technology*, pp.1-24.

[15] Park, K.C., Motai, Y. and Yoon, J.R., 2017. Acoustic fault detection technique for high-power insulators. *IEEE Transactions on Industrial Electronics*, *64*(12), pp.9699-9708.

[16] Liang, W., Zhang, L., Xu, Q. and Yan, C., 2013. Gas pipeline leakage detection based on acoustic technology. *Engineering Failure Analysis*, *31*, pp.1-7.

[17] Aymerich, F. and Staszewski, W.J., 2010. Impact damage detection in composite laminates using nonlinear acoustics. *Composites Part A: Applied Science and Manufacturing*, *41*(9), pp.1084-1092.

[18] Martinsson, J., Runefors, M., Frantzich, H., Glebe, D., McNamee, M. and Mogren, O., 2022. A novel method for smart fire detection using acoustic measurements and machine learning: proof of concept. *Fire technology*, *58*(6), pp.3385-3403.

[19] Xiong, C., Wang, Z., Huang, Y., Shi, F. and Huang, X., 2022. Smart evaluation of building fire scenario and hazard by attenuation of alarm sound field. *Journal of Building Engineering*, *51*, p.104264.

[20] Abbasi, M.Z., Wilson, P.S. and Ezekoye, O.A., 2020. Change in acoustic impulse response of a room due to a fire. *The Journal of the Acoustical Society of America*, *147*(6), pp.EL546-EL551.

[21] McLoughlin, I., Zhang, H., Xie, Z., Song, Y., Xiao, W. and Phan, H., 2017. Continuous robust sound event classification using time-frequency features and deep learning. *PloS one*, *12*(9), p.e0182309.

[22] Mesaros, A., Heittola, T. and Virtanen, T., 2016, August. TUT database for acoustic scene classification and sound event detection. In *2016 24th European Signal Processing Conference (EUSIPCO)* (pp. 1128-1132). IEEE.

[23] Takahashi, N., Gygli, M., Pfister, B. and Van Gool, L., 2016. Deep convolutional neural networks and data augmentation for acoustic event detection. *arXiv preprint arXiv:1604.07160*.

[24] Piczak, K.J., 2015, October. ESC: Dataset for environmental sound classification. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 1015-1018).

[25] Salamon, J. and Bello, J.P., 2017. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal processing letters*, *24*(3), pp.279-283.

[26] Takahashi, N., Gygli, M., Pfister, B. and Van Gool, L., 2016. Deep convolutional neural networks and data augmentation for acoustic event detection. *arXiv preprint arXiv:1604.07160*.

[27] Huang, M.X., Li, Y., Nazneen, N., Chao, A. and Zhai, S., 2021, May. Tapnet: The design, training, implementation, and applications of a multi-task learning cnn for off-screen mobile input. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-11).

[28] Li, J., Brown, C., Dzikowicz, D.J., Carey, M.G., Tam, W.C. and Huang, M.X., 2023. Towards real-time heart health monitoring in firefighting using convolutional neural networks. *Fire Safety Journal*, *140*, p.103852.

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

[30] Tam, W.C., Fu, E.Y., Li, J., Peacock, R., Reneke, P., Ngai, G., Leong, H.V., Cleary, T. and Huang, M.X., 2023. Real-time flashover prediction model for multi-compartment building structures using attention based recurrent neural networks. *Expert Systems with Applications*, *223*, p.119899.

[31] Duhamel, P. and Vetterli, M., 1990. Fast Fourier transforms: a tutorial review and a state of the art. *Signal processing*, *19*(4), pp.259-299.

[32] Logan, B., 2000, October. Mel frequency cepstral coefficients for music modeling. In *Ismir* (Vol. 270, No. 1, p. 11).

[33] Fan, L., Tam, W.C., Tong, Q., Fu, E.Y. and Liang, T., 2023. An explainable machine learning based flashover prediction model using dimension-wise class activation map. *Fire Safety Journal*, *140*, p.103849.

[34] Wang, Z., Zhang, T. and Huang, X., 2024. Explainable deep learning for image-driven fire calorimetry. *Applied Intelligence*, *54*(1), pp.1047-1062.

Figure captions

Figs. 1. Screenshot for the moment of a) breakage of safety valve (SVB) at around 331 s, b) initial gassing at around 395 s, and c) LiB ignition at around 437 s.

Fig. 2. Schematic of the experimental setup.

Figs. 3. Acoustic data of a) Experiment 30 and b) three selected events.

Figs. 4. Acoustic data of a) Experiment 12 and b) three selected events.

Figs. 5. Data distribution for a) training, b) validation, and c) testing subsets of the entire dataset that contains 2458 samples.

Fig. 6. Overall structure of the detection model.

Fig. 7. Validation loss and accuracy plot.

Figs. 8. Predictions of M14 w DA for a) Experiment 30 and b) Experiment 12 with data augmentation.

Figs. 9. Predictions of M14 wo DA for a) Experiment 30 and b) Experiment 12 without data augmentation.