# A Novel Method for Smart Fire Detection using Acoustic Measurements and Machine Learning – Proof of concept

## Abstract

Fires are a major hazard resulting in high monetary costs, personal suffering, and irreplaceable losses. The consequences of a fire can be mitigated by early detection systems which increase the potential for successful intervention. The number of false alarms in current systems can for some applications be very high, but could be reduced by increasing the reliability of the detection system by using complementary signals from multiple sensors. The current study investigates the novel use of machine learning for fire event detection based on acoustic sensor measurements. Many materials exposed to heat give rise to acoustic emissions during heating, pyrolysis and burning phases. Further, sound is generated by the heat flow associated with the flame itself. The acoustic data collected in this study is used to define an acoustic sound event detection task, and the proposed machine learning method is trained to detect the presence of a fire event based on the emitted acoustic signal. The method is able to detect the presence of fire events from the examined material types with an overall F-score of 98.4%. The method has been developed using laboratory scale tests as a proof of concept and needs further development using realistic scenarios in the future.

**Keywords:** Fire detection, artificial intelligence, machine learning, deep neural networks, acoustic emissions, sound

## 1 Introduction

Fires are a major hazard, generating direct costs for our society in the order of approximately 1-2% of the GDP in many developed countries [1]. In addition to the direct monetary cost of fires there are additional consequences in terms of personal suffering, property and environmental loss, and loss of vital societal functions for short or long periods of time. One way to mitigate the consequences of unwanted fires is to ensure their detection at an early stage, thereby increasing the potential for successful intervention.

Fire detection systems are common in most industrial facilities, assembly premises, hotels and health care facilities. A recent study, however, indicated that the number of false alarms on automatic alarm systems in Germany could be as high as 87% [2]. Similar data from Sweden indicates that false alarms could actually be as high as 97% in certain applications [3, p. 81]. Independent of the actual size of the problem, false alarms are a serious problem for the owner of a facility, for people in the building and for the fire and rescue services called to the building to respond to the fire. A false fire alarm creates unnecessary interruptions to business operations, forces people to evacuate and introduces a high unnecessary traffic risk. The reasons for false alarms vary depending on the type of detector, its application and position; but, reasons may include non-fire particles in a dirty industrial environment or produced by cooking in a kitchen (whether domestic or industrial), or steam produced in industrial or domestic situations. In essence, there are two solutions to this problem, either false alarms are stopped by organisational measures, i.e. a fire must be confirmed by a complementary means before activating the detection system to initiate a response; or the reliability of the detector is increased through a variety of technical measures. In the latter category, some effort has been made to study multi-sensor fire detection to improve the reliability of detection and reduce the number of false alarms [4, 5]. Such systems typically rely on a combination of traditional sensors and data treatment to reinforce detection reliability by confirmation of detection through several fire characteristics such as, e.g. smoke, temperature, CO-emissions and CO2-emissions (see e.g. [6], [7]). While such efforts have been successful in improving the level of detection compared to single sensor detectors [6]), they typically rely on a range of chemical (e.g. species) detection methods, heat and particle detection [8].

In recent years, papers have been published concerning the use of various types of machine learning to improve the development of algorithms to analyse these multi-sensor signals (see e.g. [9] and references therein). However, the authors have not been able to find recent papers that refer to the use of audio signals to detect fires. The closest study in the literature concerns a recent article where the authors detect the position of a fire and its characteristics using sound emitted by the detector rather than the fire, in an effort to improve data collection as input to tactical response to the fire [10]. In this application, Xiong et al. [10], assume that the fire itself has been detected by other means and the sound is produced by the alarm itself. But using acoustic signals as a way to detect fires is not addressed in the paper.

Clearly, there is a need to improve the capability for simple detectors to perform with a high level of reliability and, in the long term, this needs to be solved in a cost effective manner [11]. Two strategies can be identified in the literature to solve the issue: improve existing detectors with respect to their sensitivity by signal filters, or find alternative fire characteristics to detect an incipient fire [12]. The current study aims to investigate the novel use of machine learning for fire detection based on acoustic measurements (FORMAS Contract# 2019-00954) in an effort to improve reliability while using simple

detection methods, i.e. the focus is on using the second method with an alternative fire characteristic being applied to detect the fire. The advantage of using sound as the fire characteristic is that it can rapidly reach the detector (more rapidly than smoke dispersion or convective heat transfer through air) and it is less hindered by physical barriers such as walls. Indeed, the concept of using acoustic measurements to detect fires was first investigated in the 1990's at the National Institute of Standards and Technology (NIST) in the US [13], although acoustic flame characteristics have been investigated since the 1960's [14], albeit without reference to fire detection. In a more modern application of acoustics to combustion phenomenon, Nair [15] investigated the use of sound to identify flame blow-off. While interesting from a combustion point of view, his methodology has not been applied to the detection investigation presented in this article as there is no assumption of a steady flame to detect burning.

The initial work by Grosshandler and Jackson [13] provided proof-of-concept for using acoustic detection, but was not pursued due to difficulties with signal to noise ratio and acoustic measurement technologies which could not detect signals for large distances at that time. In the initial study, the efficacy was limited since the detection algorithm was based on hard-coded algorithms. Sound sources are often characterized only by the sound power they are emitting. Sometimes, characteristics of the frequency domain are also studied, with e.g. high-pass or low-pass filters, but it may not be possible to differentiate between vastly different sources, such as music and construction noise, with analytical or numerical methods unless the time domain is considered. When the time signal is considered, however, typically either very simple relationships are considered, e.g. counting the number of events over a threshold level, or complicated processes are studied that depend on well-defined and stable conditions. One example is the Minor Component Analysis (MCA) based method to detect signatures in the time domain presented by Kwan et al. [16].

Recently, machine learning has made great progress for many applications, due to algorithmic developments together with progress in computational capacity and the availability of large datasets with labelled data. Indeed, in a review by Naser [17], the application of machine learning and artificial intelligence in fire engineering and sciences was explored. Naser identified the use of machine learning in enhancing fire detection in domestic applications and wildland fires, but relied on traditional sensors or picture information. In no case that we have found has machine learning been applied to acoustic signals. One of the most successful approaches is deep-learning, which applies artificial neural networks (ANN) with many layers to problem solving. To date, deep learning has been used in such diverse applications as to achieve computer vision for self-driving cars, speech recognition, automatic translation, and text summarization [18]. In particular, a method called convolutional neural networks has completely redefined the state-of-the-art for processing images, video, and audio. In the area of fire safety, several

attempts have been performed using computer vision techniques based on deep learning for fire detection [19–21]. Deep learning models have the capacity to discriminate complex patterns in high-dimensional data, potentially overcoming the limitations in early approaches to sound-based fire detection mentioned above. This paper provides a proof-of-concept for using acoustic measurements together with machine learning for rapid fire detection. The aim is that this proof-of-concept will provide the foundation for more applied research into early fire detection in real fire scenarios where traditional fire detectors are prone to false alarms such as dirty industrial environments, or domestic environments with confounding issues making detection difficult. The novelty of this article is the combination of machine learning techniques with those of acoustic measurements of a simple fire. The aim is to simplify the fire scenario by using a standardised fire test methodology, i.e. the cone calorimeter (ISO 5660). This simplified application has been chosen to limit the research question to whether it is possible to use machine learning to teach a system to recognise whether there is a fire or not. Future studies will explore such questions as, which scenarios are most relevant to acoustic fire detection and whether additional acoustic complications (e.g. additional background noise) invalidate the method. When conducting this work, the authors were in agreement that there is a clear need to simplify the application in this first step and to add layers of complexity in the next step.

# 2 Theory

This section contains theory for acoustic emissions from fires and machine learning for sound event detection which are relevant to understand the contents of the paper.

## 2.1 Sound generation mechanisms for fire

Acoustic emission is an essential element of the fire detection algorithms developed in the current study. Fire detection using acoustic emission has been evaluated by Grosshandler and Braun [22] (who actually measured surface vibrations) and by Kwan et al. [16]. These initial studies have, to date, not been pursued further, particularly due to the high risk of false alarms and the problems associated with formulating threshold rules, based on analytical or signal process approaches that can function in noisy environments.

Acoustic emission is defined in ISO 2007 [23] as the range of phenomena that results in the generation of structure-borne and fluid-borne (liquid or gas) propagating waves due to the rapid release of energy from localized sources within and/or on the surface of a material. The sounds are typically either very short transient signals of a wide frequency range, or more continuous signals with narrower frequency distribution due to e.g. leaking heated fluids. There are several types of sound associated with different stages of fire development, from heating and ignition to flaming combustion.
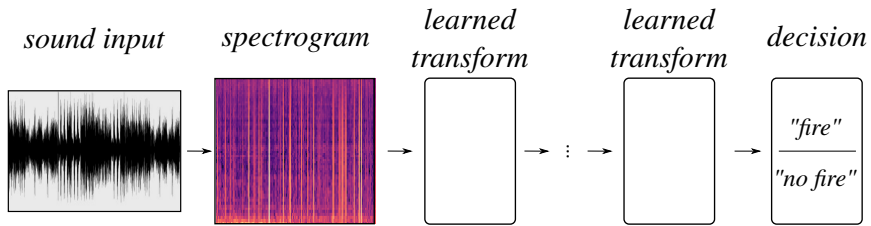
**Fig. 1**: In this work, we consider machine learning models based on deep convolutional neural networks, consisting of a series of transformations or layers. Each layer consists of a linear transformation (matrix-vector multiplication), and an element-wise nonlinearity. Convolutional layers also incorporate a convolution operation which provides spatial invariance. The input to the neural network is a sound signal which is transformed into a Mel spectrogram.

For flaming combustion, sound is generated by hydraulic instabilities and turbulence in the flame and fire plume and is typically located in the infrasound frequency range, although the resulting generated sound may be within the audible range. Detriche and Lanore [24] investigated the pulsation characteristics of small pool fires in 1980 and concluded that the signal was very sensitive to surrounding conditions, making it difficult to use analytical sound characterisation for detecting a fire. It should be noted that improved sensor technology has been developed since the 1980's, and progress in signal processing and data analysis techniques might motivate these studies to be revisited.

Both during heating and flaming combustion, sound is typically emitted by the fuel itself. Indeed, the sound that is generally associated with the moniker "fire sound", is that generated and emitted by heated material. For solid materials, the sounds originate from internal stress due to the physical decomposition and deformation of the material during the heating, pyrolysis and burning phases. A typical example is the crackling sound from burning a log of wood originating from evaporation of small pockets of trapped water in the material. Liquid fuels, as well as some thermoplastics which melt before burning (e.g. PMMA), sound can also be emitted due to boiling of the fuel.

Fires can also induce sounds not directly linked to the combustion process. One example is a paper recently published by Thompson et al. [25] where they use the sound of firebrands as they impact on a steel box to both detect the location of the flame front as well as the fire intensity at that location.

## 2.2 Machine learning for sound event detection

Machine learning techniques such as deep neural networks have revolutionized many fields, including computer vision [18]. Recently, learnings from the field of computer vision have been transferred to sound event detection. The goal of sound event detection is computerized analysis of acoustic signals for detection of sound events, i.e. what is heard and when does a specific event

occur [26]. Deep neural networks contain a sequence of simple transformations which are usually trained together end-to-end, and learn a hierarchy of representations for the input signal, in each step transforming the data into a space more suitable to solve the end task (see figure 1). Similar to two-dimensional optical images, the time-frequency representation of audio signals have successfully been modelled using a version of neural networks known as convolutional neural networks [27].

Convolutional neural networks learn to detect complex patterns in a spatially organized input. This includes the potential of spatial invariance, which allows a vision model to detect patterns at different spatial locations, and an acoustic model to detect patterns at different locations in the time-frequency domain. Compared to other machine learning approaches (including Multilayer Perceptrons), a convolutional network is relatively parameter efficient, and has a larger field-of-view. This class of models obtains state-of-the-art results in many tasks within both vision and acoustics.

# 3 Experimental methodology

This section contains a description of how the sound data from the fire experiments were produced and collected, as well as a description of the experimental setup for training and designing a deep convolutional neural network for sound event detection.

## 3.1 Fire experiments

Fire experiments were performed using the cone calorimeter [28], which is one of the most widely used instruments for ignition tests within fire science and known to develop repeatable and reproducible results. The method has been standardized in ISO 5660 and ASTM E1354 (see figure 2) but basically, the apparatus was applied as a radiative panel without any measurements of effluent gases, heat release rate or weight loss. The apparatus was slightly modified, i.e. the hood was removed to reduce the sound generated from the fan and instruments in the cone calorimeter on the recordings.

Typically, the experimental procedure was to heat up the cone, mount the sample in the sample holder and prepare the recording devices, i.e. the recorder and the microphone. The next step was to initiate the recording, place the sample holder under the pre-heated cone and open the aperture shielding the radiative cone from the sample, thus starting the experiment. The timing started as the aperture opened. The ignition process is inherently unsteady and, therefore, no distinction between the different stages in the burning process (see Section 2) has been made. Each experiment was terminated when the sample material stopped burning (see table 2). This was deemed appropriate for the current study which is a proof-of-concept, future studies should investigate possible differences in acoustic signatures for different stages. Even as the purpose with the experiment was to record sound

**Fig. 2**: The cone calorimeter setup used in the experiments.

during the preheating time, i.e. before the material ignited, the recording continued until the material stopped burning to collect also sound during the burning phase. A set of six materials or material combinations were chosen for the initial investigation. Approximately half the tests included wood (softwood (spruce), hardwood (oak) and chipboard), with the remaining containing plastic (polymethylmethacylate (PMMA), polyurethane (PUR) and a PUR/fabric combination). The choices were made to explore a range of common material and fire performance, i.e. charing and melting, see table 1. Cone calorimeter samples are typically preconditioned according to ISO 5660-1 to minimize sample variability between tests. However, in this application, variability was desirable to prevent overfitting of the model. Therefore, the samples were not preconditioned. To make sure that the signal detected is the fire event, and not the noise from the cone, the isolated sounds emitted by the cone without any sample material was also collected (see table 2).

Even as the hood over the cone was removed, there was still some background noise in the room, mainly emitted by the ventilation system in the lab. To reduce the influence of noises due to the ventilation system in the machine learning phase, the fan was arbitrarily turned on and off during some trials. Timing for when the fan should be turned on or off was sampled randomly between 30 and 50 seconds (see table 2). Further, acoustic damping using mineral wool was mounted on nearby rigid steel surfaces (see figure 2).

The sound was recorded using a Zoom H2n, with a sampling frequency of 96 kHz/24 bit, connected to an external microphone of type Earthworks Audio M23. The microphone was placed approximately 100 mm from the sample. The distance between the microphone and the sample was chosen short enough to be able to detect sound from the material decomposition and at the same time

with a safe enough distance from the radiative heat source to not damage the microphone. It is desirable to position the microphone as close to the sample as possible as the sound pressure reduces by the square of the distance. The samples were all 100 x 100 mm. Sample material, sample thickness, incident radiation and the data collected is presented in table 1.

**Table 1**: Description of the sound data recorded for the fire event class detailing the different material types, the radiation used during the heating phase, the thickness of the material, the number of trials (whole experiment, including the heating, pyrolysis and burning phases), and the total amount of recorded time for each material type.

| Sample | Radiation | Thickness | Number of trials | Total recorded time |
| --- | --- | --- | --- | --- |
| - | kW/m2 | mm | - | min |
| Oak | 35 | 45 | 3 | 33 |
| Oak | 30 | 45 | 1 | 15 |
| Oak | 35 | 10 | 4 | 22 |
| Spruce | 30 | 43 | 1 | 18 |
| Spruce | 35 | 43 | 15 | 178 |
| PMMA | 30 | 10 | 5 | 61 |
| PMMA | 35 | 10 | 1 | 8 |
| PUR | 35 | 50 | 1 | 2 |
| PUR/fabric | 35 | 50 | 1 | 5 |
| Chipboard | 35 | 10 | 3 | 19 |

All recordings where a sample material and radiation is present are considered as fire events (see table 1), and recordings without either a sample or radiation are considered as non-fire events (see table 2). These acoustic data recordings were used to train a machine learning model to distinguish between fire events and non-fire events, which is further explained in Section 3.2.

**Table 2**: Description of the sound data recorded for the non-fire event class detailing the presence of fan noise, radiation noise, the number of trials and the total recorded time.

| Fan | Radiation | Number of trials | Total recorded time |
| --- | --- | --- | --- |
| - | kW/m2 | - | min |
| On | 0 | 1 | 5 |
| Off | 0 | 1 | 5 |
| On | 35 | 3 | 17 |
| Off | 35 | 2 | 20 |
| Varying | 30 | 1 | 15 |
| Varying | 35 | 1 | 15 |

## 3.2 Machine learning for acoustic fire detection

This section presents the way the model has been trained to distinguish between fire and non-fire events on acoustic recordings of fires, and gives a description of the model architecture[1].

### 3.2.1 Training setup

The acoustic recordings of fire events and non-fire events were first split into training, validation, and test sets. The training set was used to train the model, the validation set was used to validate the model *during training*, and the test set was used to evaluate the performance of the final model. The recordings were down-sampled to 32kHz to reduce the computational cost and further split into 5 second long segments without overlap. The segments were then uniformly and independently sampled, without replacement, into the training (70%), validation (10%) and test (20%) set respectively. The resulting training, test and validation all have a class imbalance and consists of 16% to 20% non-fire events and 80% to 84% fire events.

The training data was split into batches of 16 segments each and these were used together with a loss-function and the model to compute the gradients used to update the model parameters. The model parameters were optimized using the optimization method Adam [29] which is an extension of the optimization method stochastic gradient decent. A loss function is used to guide the optimizer. Since there are two classes, fire event and non-fire event, this was modeled as a binary classification problem. Binary cross-entropy was the loss function used. An epoch of training is one iteration through the whole training dataset. The model was trained until no more improvements in the loss were observed on the validation data during the previous 100 epochs after which the model with the lowest validation loss was chosen. The training and validation loss curves are shown in figure 3, where the model has been trained for a total of 218 epochs meaning that the model with the lowest validation loss was observed at epoch 118 which is the chosen model.

### 3.2.2 Model

The state-of-the-art convolutional neural network model introduced by Kong et al. [30] was used to model the acoustic data. The architecture was designed for classification of sound events, and has been shown to transfer well between different problem domains. The architecture has 14 layers (see Table 3) and takes as input a time-frequency representation of the audio waveform. The time-frequency representation is a Mel spectrogram [31] which is a series of short-time Fourier transforms on sequences of the input data followed by a Mel filter-bank which projects them onto Mel bins. While designing a filter-bank specifically for this task may be beneficial, the development and evaluation of

---

[1]A complete description of the training setup and the model, as well as instructions on how to reproduce the main results of this study can be found in the Git-repository: https://github.com/anonymous7483/fire-event-detection-dataset/
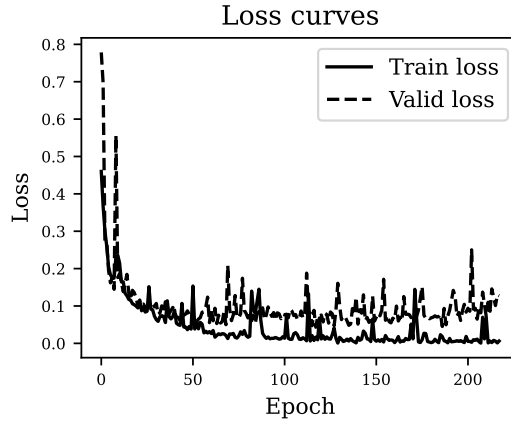
**Fig. 3**: The training and validation loss curves observed during model training.

**Table 3**: The 14 layer convolutional neural network architecture, consisting of 12 convolutional layers with a kernel size of 3 x 3 and different feature map sizes according to the table. All followed by a batch normalization layer and a rectified linear unit (ReLU) activation function. The last two layers are fully-connected layers of size 2048 and 1 with a ReLU activation and a sigmoid activation respectively.

| Model architecture | CNN14 |
|---|---|
| Input | Log-mel spectrogram, 64 Mel bins |
| Layers | (3x3 @ 64, BN, ReLU) x 2 |
| | (3x3 @ 128, BN, ReLU) x 2 |
| | (3x3 @ 256, BN, ReLU) x 2 |
| | (3x3 @ 512, BN, ReLU) x 2 |
| | Average pooling 2 x 2 |
| | (3x3 @ 1024, BN, ReLU) x 2 |
| | Average pooling 2 x 2 |
| | (3x3 @ 2048, BN, ReLU) x 2 |
| | Global average pooling |
| | FC 2048, ReLU |
| Output | FC 1, Sigmoid |

this is left for future work. In this work, the Mel filter-bank is used because of its general applicability and for being the standard choice in the machine listening literature [26].

Any audio segment which is assigned a sigmoid output score of more than a threshold $\tau$ is considered as a fire event, otherwise a non-fire event. This threshold can be adjusted, a higher threshold means that the network needs to assign a higher score for an event to be considered as a fire event, which is a way to adjust how sensitive the model is.

**Table 4**: Parameters for the Mel spectrogram.

| | |
|---|---|
| Window length | 1024 |
| Hop length | 320 |
| Window | Hanning |
| Mel bins | 64 |

### Input representation

The input to the model is a 5 second waveform with a sample rate of 32kHz, resulting in $160,000$ samples. A window of size 1024 is moved over the waveform with a hop length of 320, and a short-time Fourier transform is applied to each windowed segment of the waveform to compute the periodogram for each windowed segment. The result is a sequence of periodograms, which is called a spectrogram. The spectrogram is then processed by a Mel filter-bank, which were chosen as a set of 64 triangular filters used to map a decibel-scaled power spectrogram onto the Mel scale [31] (see table 4 for a summary of the parameters).

### Model layers

The Mel spectrogram passes through the convolutional neural network which consists of several different layers (see table 3). In the table "(3x3 @ 64, BN, ReLU) x 2" denotes a convolutional block, which consists of a convolutional layer with a kernel of size 3x3 wich outputs 64 feature maps (3x3 @ 64), followed by a batch normalization layer (BN) and a rectified linear unit (ReLU), applied twice (x2) in that order. Standard average pooling layers are used to reduce the dimensions of the representation, and finally a global average pooling layer is used to take an average over the time-dimension before applying two fully-connected (FC) neural network layers to the final representation of the input. During the training phase a dropout layer with a dropout fraction of 0.2 is applied after each convolutional block.

Dropout [32] is used to prevent over-fitting during training, which is when the model learns the training data too well, and starts performing worse on validation and test data. Batch normalization [33] is used to reduce *internal covariate shift*, and is a way to stabilize the training of the neural network and to speed up convergence.

The rectified linear unit (ReLU) is a non-linear activation function:

$$f(x) = max(x, 0), \tag{1}$$

which has become a standard activation function in the deep learning literature. Compared to e.g. the sigmoid function, the ReLU function requires little computation, and it is argued to reduce the problem of vanishing gradients.

### Output representation

A fire event is modeled as a 1 and a non-fire event as a 0 using the logistic function:

$$f(x) = 1/(1 + e^{-x}), \tag{2}$$

where $x$ is the output of the last fully-connected layer in the deep convolutional netural network.

# 4 Results

This section contains the results from the analysis of the acoustic signals collected from the different fires, and presents the evaluation results of the final sound event detection model when applied to the test data.

## 4.1 Acoustic recordings of fire events

A dataset was collected as described in section 3.1. Details of the data can be seen in table 1 and table 2.
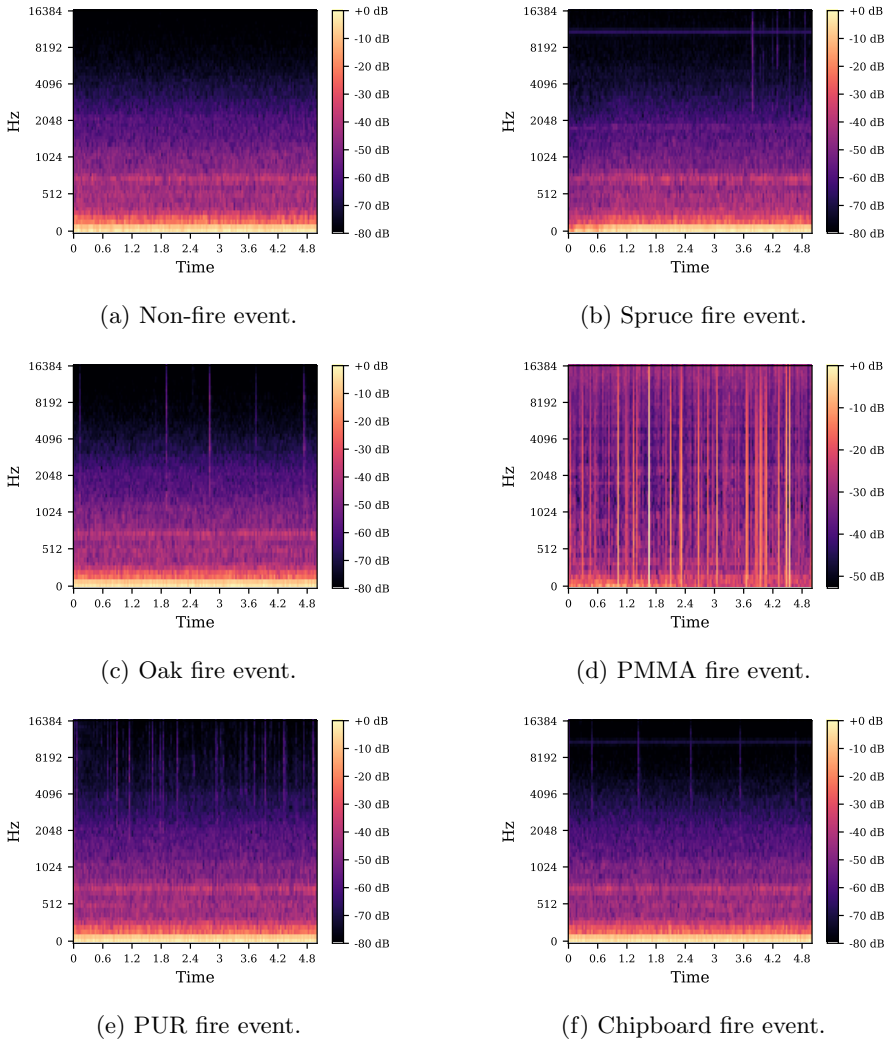
Figure 4 shows an example Mel spectrogram for a 5 second segment for each material type. These are arbitrary examples which have been chosen to provide a visual understanding of the difference between the sound events that occur for the different materials. For the human observer it is easy to distinguish PMMA fire event from a non-fire event, however, for the other materials, the distinction is not as clear. There are clear transient sounds in the recordings from all materials, and, by manual inspection of many of these Mel spectrograms, these transient sounds are the least visually prominent for the recordings of oak fire events.

## 4.2 Fire event detection using a convolutional neural network

This section presents the results from the analysis of using a deep convolutional neural network for acoustic fire event detection. All results are presented for two different values of $\tau$ (see Section 3.2) where $\tau = 0.5$ is the default choice, and $\tau = 0.97$ is chosen such that the number of false positives using the validation data is zero. The effect of $\tau$ can be seen in figure 6.

The main results which demonstrate the effectiveness of the method on the collected data are presented in table 5. The model achieves a 97.1% accuracy on the test set for the default value of $\tau$ and a precision, recall and F-score all equal to 98.4%, which means that there are equally many false positives as false negatives, in this case 14 of each. Choosing $\tau = 0.97$ means that the model becomes less sensitive towards detecting false positive fire events at the cost of becoming more sensitive towards detecting false negative events. That is, trading precision for recall. The overall performance of the model decreases, but maintains a high accuracy and F-score.

The fire event class consists of recordings of fire events from five different material types: spruce, oak, PMMA, PUR and chipboard, and the non-fire event class consists of recordings when there is no material present, i.e. the material type "none". A further analysis of the model performance on each different material type is shown in table 6. The effect of $\tau$ is apparent in this table which shows that the less sensitive model achieves a higher accuracy on

(a) Non-fire event.

(b) Spruce fire event.

(c) Oak fire event.

(d) PMMA fire event.

(e) PUR fire event.

(f) Chipboard fire event.

**Fig. 4**: Mel spectrogram visualization of the waveform collected from each different material type.

the non-fire events at the cost of achieving a lower accuracy for in particular the oak material, but also slightly lower for spruce and PMMA.

In figure 5 the accuracy of the model on fire events (lower right) and non-fire events (upper left) is presented, as well as the false positive (upper right) and false negative (lower left) rate, for different values of $\tau$.

**Table 5**: The accuracy, precision, recall and F-score for the CNN14 model on the test set with two different threshold values. A fire event is considered as the positive class, and a non-fire event is considered as the negative class.

| Metric | $\tau = 0.5$ | $\tau = 0.97$ |
|---|---|---|
| Accuracy | 97.3 % | 95.1 % |
| Precision | 98.4 % | 99.6 % |
| Recall | 98.4 % | 94.5 % |
| F-score | 98.4 % | 97.0 % |

**Table 6**: Model accuracy for each respective material in the test set, with $\tau = 0.5$ and $\tau = 0.97$.

| Material | Accuracy ($\tau = 0.5$) | Accuracy ($\tau = 0.97$) |
|---|---|---|
| None | 91.6 % | 98.2 % |
| Spruce | 99.8 % | 98.1% |
| Oak | 92.9 % | 78.2 % |
| PMMA | 99.4 % | 98.9 % |
| PUR | 100 % | 100 % |
| Chipboard | 100 % | 100 % |



(a) $\tau = 0.5$                    (b) $\tau = 0.97$
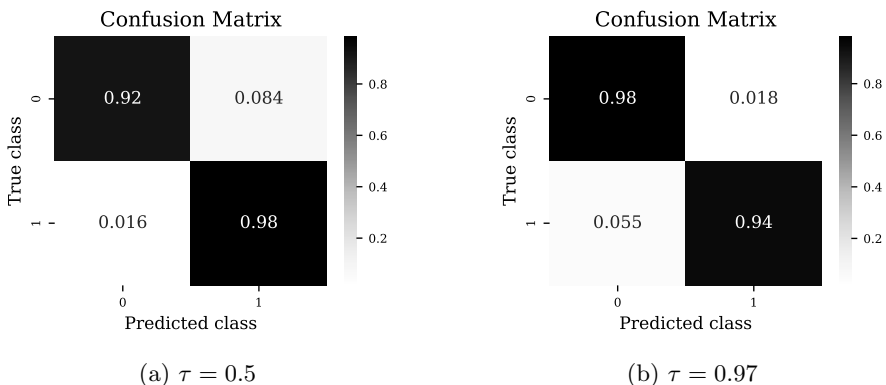
**Fig. 5**: The normalized confusion matrix for $\tau = 0.5$ (a) and $\tau = 0.97$ (b).

# 5 Discussion

The current study presents a setup and method for data collection of acoustic signals from fire events. The collected acoustic signals are used to define a classification task for fire event detection. A convolutional neural network is used to model the acoustic signal and to detect the fire event. These fire events are shown to be detectable with an accuracy of 97.3%, a precision of 98.4%, a recall of 98.4%, and an F-score of 98.4% when the threshold $\tau$ is set to 0.5.
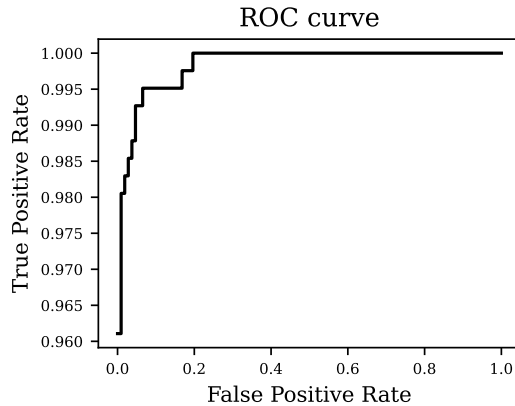
**Fig. 6**: The effect of different threshold values on the true positive and false positive rate.

That is, the fire events, as defined in this study, are shown to be detectable from the acoustic signal using a convolutional neural network.

Note that the class imbalance in this dataset does not reflect what is expected in most real settings where non-fire events would be expected to greatly outnumber fire events. The presented accuracy of the model should therefore be read with that in mind. The F-score and ROC curve are presented as a complement which are suitable metrics for imbalanced datasets.

The accuracy of the fire event detector varies depending on the material being exposed to the heating condition. The materials that give rise to a very distinct acoustic signal, such as PMMA, are detectable with very high accuracy, and the materials that give rise to a less distinct acoustic signal, such as oak, are harder to detect. Of the wood samples tested, spruce is detected with the highest accuracy and it can be hypothesised that this is due to the more pronounced crackling sound associated with spruce compared to oak. However, the sound produced by the flame and fire plume during the combustion phase could also have an effect. Also, the external conditions like initial temperature and moisture content may also have an influence on the acoustic characteristics, especially for wooden based samples. The sensitivity of the model to variations in temperature and moisture was somewhat decreased by using non-preconditioned samples, and thus a variability in this respect in the training set. However, the ability of the method to be generalized to other materials and conditions than those present in this study is not known, and to take this work from a proof of concept stage to a realistic task, more materials and fire scenarios are needed in the data set.

It should also be noted that the heating conditions used in this study are not necessarily representative of how most actual fire starts, but were chosen as a way to isolate the acoustic fire event signal of interest from other potential acoustic signals to demonstrate that it is feasible to use acoustic measurements

for fire detection. The potential influence of heating conditions is not known at this time although efforts were made to compensate for sounds emitted from the heating cone. A benefit of the chosen experimental setup is that it is well known in the fire community and known to deliver results that are repeatable and reproducible.

The strength of a data driven method is that it can be adapted to a new environment, either by training the model using data collected from such an environment, or using data which has been augmented to resemble such an environment. A limitation in the data collection setup in this study is that there was not much variance in the acoustic environment. In a real setting there may be other noise sources present such as talking humans and driving vehicles, and the impulse-response of the acoustic room may also vary depending on e.g. the size of the room and the material of the walls.

To make the model more robust against varying acoustic environments the training and test data need to capture this variance. A way to mitigate the need for such costly data collection efforts is to augment the already existing data with other noise sources by simply mixing multiple acoustic signals together. To emulate different acoustic rooms the impulse-response of such environments could also be taken into account when mixing the signals.

The distance between the fire and the microphone will have an effect on the performance of the system. In this study, we collected data where the sound source was 100 mm from the microphone (see Section 3.1). A number of sources of noise is present in this data; most notably ventilation and electrical interference. At a greater distance, the increased signal-to-noise ratio will make fire prediction harder but we hypothesize that the solution will have potential if the training data is extended to cover this variance through further collection or data augmentation. We leave it to future work to study this effect in detail.

Another promising way of reducing the need for extensive data collection is transfer learning. The neural network architecture used in this study has been developed and shown to transfer well between different acoustic tasks, and pre-training the network on similar acoustic data is an interesting way forward. Transfer learning and data augmentation could therefore be two important ways forward to take this a proof of concept to a method applicable in a more realistic setting.

The data collected in this study, together with the annotations, have been made publicly available to facilitate further research on fire event detection using acoustic signals. Instructions on how to download the data can be found in the supplementary material.

Interesting future work would be to treat this as a regression problem and, e.g., study if it is possible to predict more detailed characteristics of the flame such as flame size or heat release rate from the acoustic signal during the kindling phase, or the time until and after the kindling phase.

# 6 Conclusions

This study investigates the use of acoustic sensors for early fire detection. Microphones are a relatively inexpensive form of sensor, and using the acoustics from a fire event as a complementary signal in current fire event detection methods can make them more robust and reliable. The results presented shows that the acoustic signal from a fire event can be used to detect fires in the setting proposed in this study. The acoustic vibrations of the materials exposed to heat are used to train a machine learning method to detect such vibrations. The results show that the machine learning method can detect fire events from measurements of the acoustic signals being emitted from the materials when heated. The analysis suggests that performance of the convolutional neural network varies depending on the material which is being exposed to the heating condition.

The proposed method provides proof-of-concept only and further research is needed to investigate, e.g. the impact of different acoustic environments and different materials on the predictive qualities of the method. Transfer learning, domain adaptation, and data augmentation are suggested as potential methods for further investigation.

**Supplementary information.** All the raw data used in this study can be found at the following Git-repository: https://github.com/anonymous7483/fire-event-detection-dataset. The repository contains instructions on how to download and pre-process the data, and how to train and evaluate the machine learning model presented in this study on the data.

# References

[1] McNamee, M., Meacham, B., van Hees, P., Bisby, L., Chow, W.K., Coppalle, A., Dobashi, R., Dlugogorski, B., Fahy, R., Fleischmann, C., Floyd, J., Galea, E.R., Gollner, M., Hakkarainen, T., Hamins, A., Hu, L., Johnson, P., Karlsson, B., Merci, B., Ohmiya, Y., Rein, G., Trouvé, A., Wang, Y., Weckman, B.: IAFSS agenda 2030 for a fire safe world. Fire Safety Journal **110** (2019). https://doi.org/10.1016/j.firesaf.2019.102889

[2] Festag, S.: False alarm ratio of fire detection and fire alarm systems in Germany - A meta analysis. Fire Safety Journal **79**, 119–126 (2016). https://doi.org/10.1016/j.firesaf.2015.11.010

[3] Hjort, B.: Automatiskt brandlarm - onödiga larm. Technical report, Räddningsverket, Karlstad (2001)

[4] To, H., Fong, N.: Investigation of the performance and improvement of optical smoke detectors. Procedia Engineering **62**, 985–993 (2013). https://doi.org/10.1016/j.proeng.2013.08.152

[5] Chen, X., Bu, L.: Research of fire detection method based on multi-sensor data fusion. 2010 International Conference on Computational Intelligence and Software Engineering, CiSE 2010 (2010). https://doi.org/10.1109/CISE.2010.5677271

[6] Milke, J.A., Hulcher, M.E., Worrell, C.L., Gottuk, D.T., Williams, F.W.: Investigation of multi-sensor algorithms for fire detection. Fire Technology **39**(4), 363–382 (2003). https://doi.org/10.1023/A:1025378100781

[7] Davis, W.D., Cleary, T., Donnelly, M., Hellerman, S.: Using sensor signals to analyze fires. Fire Technology **39**(4), 295–308 (2003). https://doi.org/10.1023/A:1025322015802

[8] Baek, J., Alhindi, T.J., Jeong, Y.-S., Jeong, M.K., Seo, S., Kang, J., Heo, Y.: Intelligent Multi-Sensor Detection System for Monitoring Indoor Building Fires. IEEE Sensors Journal **21**(24) (2021). https://doi.org/10.1109/JSEN.2021.3124266

[9] Yu, C., Fang, J., Wang, J., Zhang, Y.: Video Fire Smoke Detection Using Motion and Color Features. Fire Technology **46**(3) (2010). https://doi.org/10.1007/s10694-009-0110-z

[10] Xiong, C., Wang, Z., Huang, Y., Shi, F., Huang, X.: Smart evaluation of building fire scenario and hazard by attenuation of alarm sound field. Journal of Building Engineering **51** (2022). https://doi.org/10.1016/j.jobe.2022.104264

[11] Khan, F., Xu, Z., Sun, J., Khan, F.M., Ahmed, A., Zhao, Y.: Recent Advances in Sensors for Fire Detection. Sensors **22**(9) (2022). https://doi.org/10.3390/s22093310

[12] Fonollosa, J., Solórzano, A., Marco, S.: Chemical Sensor Systems and Associated Algorithms for Fire Detection: A Review. Sensors **18**(2) (2018). https://doi.org/10.3390/s18020553

[13] Grosshandler, W., Jackson, M.: Acoustic emission of structural materials exposed to open flames. Fire Safety Journal **22**(3), 209–228 (1994). https://doi.org/10.1016/0379-7112(94)90012-4

[14] Thomas, A., Williams, G.T.: Flame Noise: Sound Emission from Spark-Ignited Bubbles of Combustible Gas. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences **294**(1439), 449–466 (1966)

[15] Nair, S.: Acoustic Characterization of Flame Blowout Phenomenon. PhD thesis, Georgia Institute of Technology (2006)

[16] Kwan, C., Zhang, X., Xu, R.: Early fire detection using acoustic emissions. IFAC Proceedings Volumes (IFAC-PapersOnline) **36**(5), 351–355 (2003). https://doi.org/10.1016/S1474-6670(17)36516-3

[17] Naser, M.Z.: Mechanistically Informed Machine Learning and Artificial Intelligence in Fire Engineering and Sciences. Fire Technology **57**(6) (2021). https://doi.org/10.1007/s10694-020-01069-8

[18] Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature14539

[19] Muhammad, K., Ahmad, J., Baik, S.W.: Early fire detection using convolutional neural networks during surveillance for effective disaster management. Neurocomput. **288**(C), 30–42 (2018). https://doi.org/10.1016/j.neucom.2017.04.083

[20] Dung, N.M., Ro, S.: Algorithm for fire detection using a camera surveillance system. In: Proceedings of the 2018 International Conference on Image and Graphics Processing. ICIGP 2018, pp. 38–42. Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3191442.3191450. https://doi.org/10.1145/3191442.3191450

[21] Lin, G., Zhang, Y., Xu, G., Zhang, Q.: Smoke detection on video sequences using 3d convolutional neural networks. Fire Technology, 1–21 (2019)

[22] Grosshandler, W., Braun, E. Fire Safety Science – Proceedings of the fourth international symposium, 773–784 (2019)

[23] ISO-22096:2007: Condition monitoring and diagnostics of machines - acoustic emission. Standard, International Organization for Standardization, Geneva, CH (2007)

[24] Detriche, P., Lanore, J.C.: An acoustic study of pulsation characteristics of fires. Fire Technology **16**(3), 204–211 (1980). https://doi.org/10.1007/BF02476759

[25] Thompson, D.K., Yip, D.A., Koo, E., Linn, R., Marshall, R. G.and Refai, Schroeder, D.: Quantifying Firebrand Production and Transport Using the Acoustic Analysis of In-Fire Cameras. Fire Technology (2022). https://doi.org/10.1007/s10694-021-01194-y

[26] Mesaros, A., Heittola, T., Virtanen, T., Plumbley, M.D.: Sound Event Detection: A tutorial. IEEE Signal Processing Magazine **38**(5), 67–83 (2021) arXiv:2107.05463. https://doi.org/10.1109/msp.2021.3090678

[27] LeCun, Y., Bengio, Y.: Convolutional Networks for Images, Speech, and Time Series, pp. 255–258. MIT Press, Cambridge, MA, USA (1998)

[28] Babrauskas, V.: Development of the Cone Calorimeter: A Bench-Scale Heat Release Rate Apparatus Based on Oxygen Consumption. NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD (1982)

[29] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015). http://arxiv.org/abs/1412.6980

[30] Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W., Plumbley, M.D.: PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition. IEEE/ACM Transactions on Audio Speech and Language Processing **28**(1), 2880–2894 (2020) arXiv:1912.10211. https://doi.org/10.1109/TASLP.2020.3030497

[31] Stevens, S.S., Volkmann, J.E., Newman, E.B.: A scale for the measurement of the psychological magnitude pitch. Journal of the Acoustical Society of America **8**, 185–190 (1937)

[32] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research **15**(56), 1929–1958 (2014)

[33] Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Bach, F., Blei, D. (eds.) Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 37, pp. 448–456. PMLR, Lille, France (2015). https://proceedings.mlr.press/v37/ioffe15.html