www.cpwr.com • www.elcosh.org

CPWR

THE CENTER FOR CONSTRUCTION
RESEARCH AND TRAINING

Intelligent Hearing Protection for Construction Workers Exposed to Hazardous Noise

Tuyen Le, Ph.D.
Kehinde Elelu

Clemson University

March 2022

8484 Georgia Avenue
Suite 1000
Silver Spring, MD 20910

PHONE: 301.578.8500
FAX: 301.578.8572

©2022, CPWR-The Center for Construction Research and Training. All rights reserved.
CPWR is the research and training arm of NABTU. Production of this document was supported by cooperative agreement OH 009762 from the National Institute for Occupational Safety and Health (NIOSH). The contents are solely the responsibility of the authors and do not necessarily represent the official views of NIOSH.

Report #20-3-PS

Abstract

The ability to hear safety cues while wearing hearing protection equipment (HPE) is critical to preventing injuries and deaths on construction job sites. The goal of this project is to improve auditory situational awareness of construction workers exposed to loud noise by investigating a new hearing protection technology that uses artificial intelligence (AI) to amplify safety-critical sounds of collision hazards while greatly attenuating ambient noise. This Small Study focused on developing a signal processing model to help workers wearing HPE improve their audible sense of mobile equipment. This study included three phases: (a) collecting audio data of construction equipment, (b) developing a novel audio-based machine learning model for automated detection of collision hazards to be integrated into intelligent hearing protection devices, and (c) conducting field experiments to investigate the system's efficiency and latency. The outcomes showed that the proposed model detects equipment correctly gave workers timely notifications of hazardous situations.

Key Findings

The key results of this study include:

- The machine learning models trained with a Convolutional Neural Network (CNN) yield reliable collision hazard predictions, with an accuracy of 88% in detecting sounds related to collision hazards when the signals are not buried in background noises. Accuracy remained at that level in loud-noise situations when the signal-to-noise ratio remains above 10db.
- The study developed a mobile application implementing the CNN model and conducted two sets of experiments, in a controlled environment and on a construction site. The results showed that the mobile application yields a high detection accuracy, particularly for equipment with unique sound patterns.

Table of Contents

Abstract i
Key Findings i
Introduction 1
Objectives 1
Methods 1
Accomplishments and results 6
Future funding plans 12
Presentations, Publications, and Dissemination Plan 12
References 12

Introduction

According to the Occupational Safety and Health Administration (OSHA), construction has an unusually high annual fatality rate compared to other industrial sectors in the U.S. (Hinze et al., 2011). Struck-by-vehicle incidents are a leading cause of construction-related deaths (Samantha et al., 2021), mainly due to the proximity between workers and heavy mobile equipment on job sites (Marks et al., 2013). Previous studies reported that the critical factor leading to collisions was the decline in auditory situational awareness of construction workers (Morata et al. 2005) and the complicated nature of construction noises (Vinnik et al. 2011). Therefore, a novel audio-based technique that can augment the audible sense of workers is crucial to improving safety performance.

Advanced computational techniques in auditory signal processing are highly applicable for collision hazard detection in construction as mobile construction equipment often produces unique sound patterns while performing certain activities (Cheng et al. 2017, Cheng et al., 2016). However, acoustic events are typically complicated by heterogeneous sound types generated from diverse equipment operations, including static equipment and hand tools (Lee et al. 2020, Xie et al., 2019). Therefore, it is useful to distinguish between acoustic events of mobile equipment, which may produce collision hazards, and acoustic events of stationary equipment. The auditory surveillance of vehicles that are potential causes of struck-by incidents would significantly improve construction safety.

However, sound sensing for safety in construction has received little attention from the academic community. Most studies focused on tracking construction equipment have focused on reducing operating costs and identifying working and operation activities (Cheng et al. 2017, Cheng et al. 2016, Sabillon et al. 2018, Sherafat et al. 2018). No previous studies have been designed to help the workers recognize important signals buried in background noises. To address this gap, we propose a novel audio-based machine learning model for the automated detection of collision hazards at construction sites. The study has two primary contributions: (1) a new labeled dataset of normal and abnormal sound events relating to collision hazards at the job site, and (2) a Convolutional Neural Network (CNN)-based sound processing model for automated detection of collision hazards.

Objectives

This study aimed to determine whether the proposed technology improves workers' safety by augmenting their ability to hear important sounds related to collision hazards. Its goals include:

1. Identifying and characterizing distinctive features of acoustic safety cues associated with abnormal equipment-related situations that require quick and effective responses from construction workers.
2. Developing an AI technique for automated recognition of auditory events over complex and ambiguous ambient noises.
3. Developing a proof-of-concept prototype of intelligent hearing protection equipment that can recognize critical sounds and alert users to potential hazards.

Methods

To enable automated detection of potential collisions associated with mobile equipment on the construction site, we developed an innovative framework using a supervised deep learning approach for distinguishing critical sound and background noises. The overall process included three main steps: (1) collecting and labeling acoustic signals as abnormal and normal types, which were mixed at different signal-to-noise ratios for testing purposes,

(2) extracting acoustic features using the Fast Fourier Transform (FFT) function, and (3) training a CNN model using the labeled data to detect acoustic events. Fig. 1 presents an overview of this study.

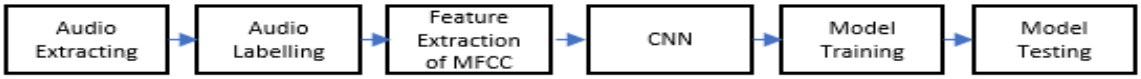

Figure 1. Framework for audio-based classification

Dataset preparation

The data preparation stage defined the set of events the system should recognize for the scope of the study, and the audio files of the dataset prepared for this research included two sources: 1) audiotapes extracted from videos downloaded from publicly available repositories and 2) sound recorded from construction sites of our industry partners. Since sounds are an essential indicator of dangerous situations requiring quick safety responses, collected sound events were manually labeled as abnormal and normal. Equipment in motion is a primary source of construction hazards and should therefore be considered to be producing abnormal sounds, while the sounds of stationary equipment were tagged as normal (see Table 1). There were 180 selected audio files in the abnormal group (20 for each of nine types of mobile equipment) and 140 selected audio files in the normal group (20 for each of seven types of stationary equipment). The duration in seconds of the audio files in each subset of the abnormal class and of the normal class are summarized in Table 1.

Table 1. Number of original examples in each subset of data

Abnormal group (Mobile equipment)			Normal group (Stationary equipment)		
Type	**Total duration(s)**	**Count**	**Type**	**Total duration(s)**	**Count**
Bulldozer	60	20	Concrete pumper	60	20
Compactor roller	60	20	Hammer	60	20
Crane	60	20	Pile driver	60	20
Excavator	60	20	Pneumatic breaker	60	20
Front end loader	60	20	Pneumatic tamper	60	20
Forklift	60	20	Saw	60	20
Grader	60	20	Steel welding	60	20
Scraper	60	20			
Water truck	60	20			
Total	540	180		420	140

The audio files were selected and recorded in high quality, avoiding noisy backgrounds, and converted into the .wav format at 16 kHz sampling rate, 16-bit depth, and mono channel. To generate audio examples that include concurrent sounds for testing purposes, the abnormal signals (mobile equipment sounds) were mixed with the normal noises (stationary equipment sounds) at different signal-to-noise ratios (SNRs). SNR represents how large the signal level is compared to the noise level, and the unit is in dB (decibels). A signal is defined as an abnormal sound that needs to be detected while noise refers to unimportant sounds of stationary equipment. The higher the SNR is, the higher the signal's amplitude is relative to that of the noise. The SNR can be calculated by the following formula:

$$SNR_{dB} = 20log_{10}\frac{A_{signal}}{A_{noise}} \qquad (1)$$

This sound mixing process generates a new dataset of 44,800 audio files, each of which is a mixture of two distinguishable equipment types: mobile equipment mixed with stationary equipment, or stationary equipment mixed with stationary equipment, as shown in Figure 3. The new audio files that include one sound from the mobile equipment group are considered abnormal, such as an excavator mixed with a hammer. The audio files do not classify any sound from mobile equipment as normal. The mixtures were created at different SNRs (-10dB, -5dB, 0dB, 5dB and 10dB) which was used for testing.

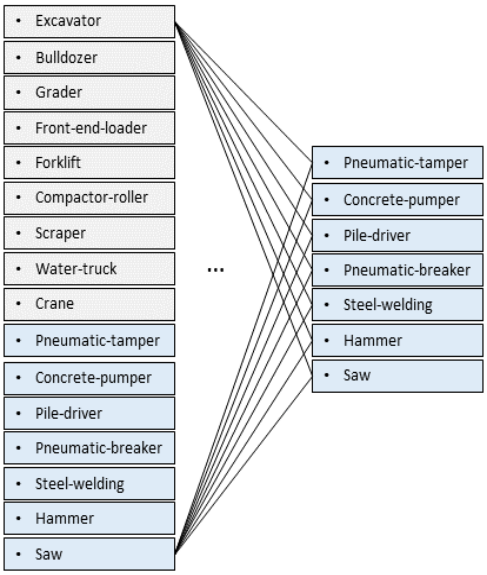

Figure 2. Diagram of sample audio file mixtures.

Feature extraction

The audio files in the dataset were used for extracting Mel-Frequency Cepstral Coefficients (MFCCs) features, the most commonly used acoustic features in signal processing (Cowling et al. 2003, Eronen et al. 2006). The extraction of MFCC involves the following three steps:

a) Framing and Windowing: In this step, the original audio files were segmented into smaller frames with an equal length of 25 milliseconds. Windowing is a standard procedure performed before spectrum calculation to minimize spectral leakage and increase spectrum sensitivity, which is unavoidable when dividing the data frame of the audio signal and introducing discontinuities at the frame border (Wieczorkowska et al, 2018). The window size should be small and large enough so that enough power spectral within each window can be obtained. The most popular window length is 25 milliseconds with 10-millisecond overlap (Alamdari et al. 2017). This window size generates 400 samples with a sampling rate of 16 kHz. The Hanning window is employed in this step to eliminate the edge effects caused by framing (Zhang et al. 2018). As a result of windowing, the original audio signal values are tapered to zero at both ends of the frame (Wieczorkowska et al, 2018).

b) Fast Fourier Transform: We used Fast Fourier Transform to convert the input signals from the spatial domain to the frequency domain. The audio is computed with a 512-point FFT using the Hanning window function (Cheung et al. 2010, Wirz et al. 2010). Due to the high-power spectral density, it is necessary to down-sample the audio at an appropriate sampling rate. Frequencies higher than half of the sampling rate, such as the Nyquist frequency, will affect the samples in a way that is misinterpreted by the interpolation process. Hence,

choosing a reasonable sampling rate would achieve the best results (Tarzia et al. 2011, Janjua et al. 2019). Audio sound is typically recorded at a rate of 44,100 kHz; in this case, the sound is down-sampled to 16 kHz since our examination of the data showed that most of the frequencies are under 8 kHz.

c) Mel-Filter Bank and Inverse Fourier Transform: The magnitude spectrum of the frequency domain is fed into Mel-filter banks. Each filter has a center frequency called the filter bank energies. This compression operation makes the acoustic features match more closely to what humans hear and produces the log of the power of the spectrum energy at each of the Mel-frequencies. In the following step, the Discrete Cosine Transform (DCT) is applied to filter bank energies. The output coefficients of DCT are called Mel Frequency Cepstral Coefficients (MFCCs). MFCC is the feature most commonly used in sound classification and detection (Cowling et al. 2003, Eronen et al. 2006). It is worth noting that our work considers MFCC coefficients in a range between 10 to 20. This is because the higher MFCC coefficients represent fast changes in the filter bank energies, and it turns out that these rapid changes degrade sound classification performance. The MFCCs extracted from the sound signal are stored as an array of values which are used as the input data when training the classification model.

Model development

We used CNN to train signal processing models that classify abnormal and normal sounds. The training process is illustrated in Figure 3 and discussed in the following subsections.

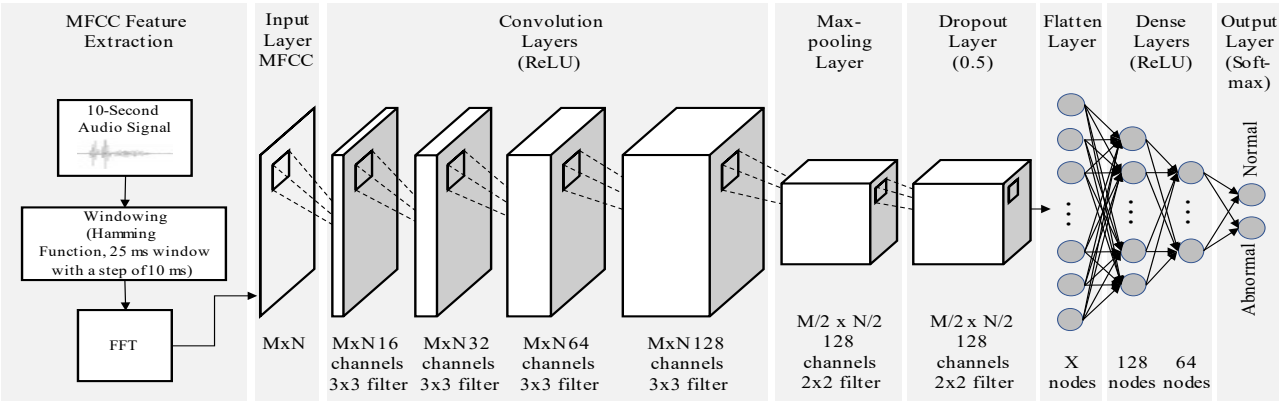

Figure 3. Audio signal processing process

CNN model

After the feature extraction was completed, the CNN model was developed for sound detection with the array of the MFCC values as the input. The size of MFCC values is M x N, where M is the number of frames and N is the number of MFCCs. The deep CNN architecture employed in this study comprises four convolutional layers with different number of channels as depicted in Figure 3, followed by a max-pooling layer, a dropout layer, a flatten layer, and two fully functional dense layers connected layers to get the output. The activation function used for convolutional and dense layers is the Rectified Linear Unit, most used in deep learning models. The function returns zero if it receives any negative input, but it returns the same value for any positive value. The SoftMax activation function is applied to the output layer. The output layer includes the predicted labels (normal or abnormal) of the input audio files.

We trained the CNN model by using 80% of the samples for training and the remaining 20% for testing. The training procedure was stopped after 20 epochs. In the baseline model, the authors initially considered 20

MFCCs, 20 filters in the filter bank, and the window size of 500ms. Then, CNN models were trained with some modification of parameters. Finally, the best value of the number of MFCCs, number of filters, and window size was applied to run the CNN model. Each model was subsequently tested on five test sets with five different SNR value (-10dB, -5dB, 0dB, 5dB, and 10dB).

Overfitting of the model was one of the errors encountered during the development of CNN models. We implemented the ensembles of trained models with different settings to address this issue. To reduce the computational requirements for the ensemble training we used dropout (a regularization technique) to randomly drop out some nodes in the neural network. This means that their contribution to the activation of downstream neurons was temporally removed on the forward pass, and any weight updates were not applied to the neuron on the backward pass. The process made use of a probability of 0.5. Even though dropout did not mean accuracy will increase, it helped prevent the most common error in CNN, overfitting of the model. The combination of ensembles and dropout is a well-known computationally cheap and remarkably effective method to reduce overfitting while improving generalization error in deep neural networks.

Evaluation

This study used a ten-fold cross-validation approach to avoid the randomness of selected validation examples when measuring the prediction performance. The training and validation data set was obtained by dividing the original dataset into ten mutually exclusive folds of data (see Figure 4). The data folds were selected in Stratified k-fold cross-validation so that each contains the same number of abnormal sound examples. Each fold was used once to validate the performance, and the remaining nine folds were used for training, which obtained ten independent performance values. This procedure was repeated ten times by changing the remaining folds, and ten prediction performances were generated. The performance of the prediction model was obtained by the average predictive results of the ten folds.

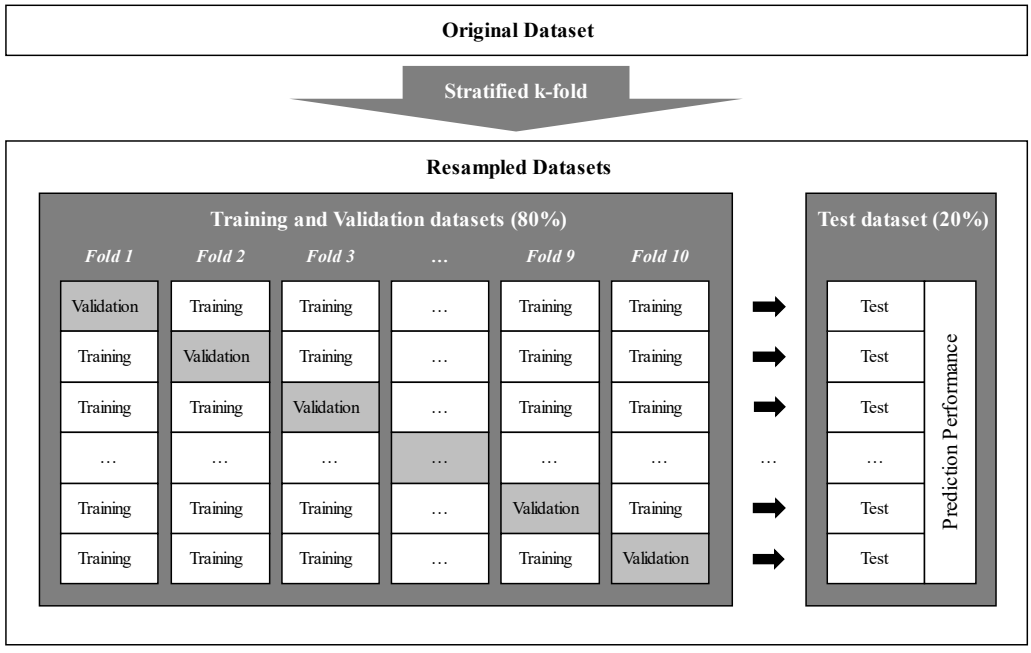

Figure 4. Ten-fold cross-validation

To measure the performance of the classification techniques, four different metrics including accuracy, precision, recall, and F1-score, were used. Accuracy was used to evaluate the sound detection performance, as shown in Equation 2. The accuracy metric is determined based on the following figures: true positive (TP), true negative (TN), false positive (FP), and false negative (FN), where TPs are the number of audio files labeled correctly as "abnormal," FPs are the number of audio files labeled incorrectly as "abnormal," TNs are the number of audio files labeled correctly as

"normal," and FNs are the number of audio files labeled incorrectly as "normal." In plain language, accuracy is the number of correct predictions divided by the total number of predictions. The accuracy reaches its best at 1 and worst at 0.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (2)$$

Precision measures how many of the true points predicted are actually true, whereas recall measures the rate of how many true points are correctly predicted, respectively expressed in Equations (3) and (4). Both precision and recall are desirable, but a trade-off between the two measures may be needed, as they can be negatively correlated. The F1-score is a combined measure of precision and recall as shown in Equation (5). The F1-score reaches its best at 1 and worst at 0.

$$Precision = \frac{TP}{TP + FP} \qquad (3)$$

$$recall = \frac{TP}{TP + FN} \qquad (4)$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \qquad (5)$$

Accomplishments and results

The CNN models were trained using Python programming language on Clemson Palmetto supercomputer clusters equipped with a CPU @ 40 GHz, V100 with NVLink GPU model, and 4 GPUs per chunk. We then developed a prototype on a mobile device that implements the trained CNN models to support automated detection of collision hazard sounds. We tested the device on both hypothetical and real construction sites.

Computational performance of the CNN models

We trained various CNN models using different neural network settings. The performance of the best model tested on each of the five test sets is summarized in Table 2. Overall, the results showed that the performance scores increased when there was less background noise in the audio files. When being tested on the dataset without overlapping sounds, the model achieved an accuracy of 87.98%. This figure dropped to 85.17% when background noise sounds, were added to the clean signals at 10db SNR. The model's performance became relatively poor when noises were significantly louder than important signals. The accuracy of the models on the -10db SNR and -5db SNR were 50.63% and 56.85% respectively.

Table 2. Comparison of model performance (frame size = 0.1s)

	Performance achieved on each test set					
Metrics	-10dB	-5dB	0dB	5dB	10dB	No mixture
Accuracy	0.5063	0.5685	0.6760	0.7785	0.8517	0.8798
Precision	0.5844	0.6466	0.7218	0.7911	0.8498	0.8779
Recall	0.5492	0.6069	0.6979	0.7920	0.8551	0.8858
F1-score	0.4697	0.5508	0.6716	0.7785	0.8507	0.8789

Field tests and experimental setup

In order to validate the applicability of the developed model in real construction sites, we built a mobile application for Android devices using TensorFlow Lite framework (a lightweight framework for training machine learning models). This mobile application provides users with alerts of the occurrence of mobile equipment, along with the probability that the detection is correct. We first converted the saved CNN model (meta graph) with the highest accuracy using a TensorFlow Lite Converter to a file format of protobuf (.pb) –

which contains the graph definition as well as the weights of the model into TensorFlow Lite (.tflite) – that enables on-device machine learning on Internet of Thing devices. During the conversion process from a TensorFlow model to a TensorFlow Lite, the size of the file is reduced, and there is a possibility of further reducing the file size, although there may be a trade-off in execution speed of the model. The CNN models, which were trained using the TensorFlow Lite framework, were then converted to the Android Studio ML Model Binding format. This allows the models to be compatible to Android devices. This metadata of the Android Studio ML Model Binding files includes the details of the trained models along with other descriptions required to deploy the model. The mobile application implementing the trained models was built using Android Studio 4.2. Fig 5 shows the Android application and loudness recorder interface, respectively.

Two separate experiments were conducted to evaluate the efficacy of the Android application in detecting the proximity of mobile equipment. The first experiment (including sixteen test signals) was carried out in the laboratory, while the second (including four test signals) was done on the field. We repeated each test three times to avoid measurement errors. For each of the experiments, the following outputs were recorded: type of equipment detected, the probability that the detection is correct, and the time it took to detect the signal. The loudness of the sound while experimenting ranges between 72 and 80db, measured using a Decibel Meter iOS application.

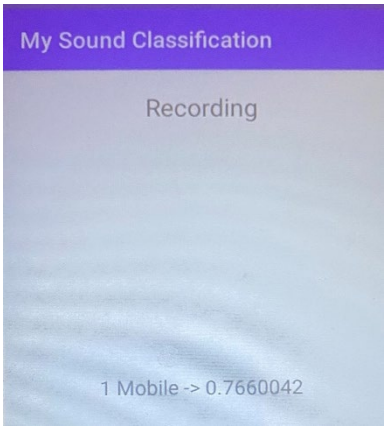

Figure 5. Sound classification android application

Figure 6 shows the experimental setting for the experiments in the laboratory. A sound source (providing 16 types of sound) generated from a computer speaker was placed four meters away from the mobile device. In the field experiments, a site engineer was asked to carry the mobile device with the mobile application installed and stand 10 meters and 20 meters from the equipment. He recorded the average time and probability of hazard detection for three samples of each of the four types of equipment. The variation in distance checked the impact of distance on the model, because equipment closer d to a worker signifies more danger. Fig 7 and 8 shows the setup for testing the model on the construction site.

Figure 6. Experiment setup in the controlled environment: the loudness detector is on the left, the computer and the mobile device on the right)

Figure 7. Sound detection experiments with an excavator and a front-end loader

Figure 8. Sound detection experiments with a hammer

Testing results

Table 3 below shows the result from the set of experiments conducted in the controlled environment. The acoustic sensing application monitored equipment classification and measured the time required to detect the auditory signal of mobile equipment. The application typically took less than 10 seconds to generate an alert. Yet it is difficult to confirm whether this latency is sufficient for real-time hazard detection, as it depends on many job site factors, such as the speed and the direction of a target mobile vehicle as well as the presence of barriers between the worker and the equipment. It would be ideal to reduce the delay, which would allow the worker more time for responsive safety actions. Shortening the delay seems to be challenging, however, at complex workplaces with excessive background noise. Because of those complexities, our model requires significant computational power for training a large amount of real-life data. Future work is needed that implements advanced pre-processing algorithms (i.e., de-noising) with lower computational requirements, which would improve the overall performance of the proposed system, particularly reducing the detection duration.

Another key observation from the field tests was that equipment working with concrete in motion tends to create significant noise, thus resulting in a lower probability value, and it requires a longer detection time than equipment like forklifts and compactor rollers. This complication can also be seen with equipment like front-end loaders: the sound detection has a relatively low probability due to the occurrence of background noise from the granite being picked up, which also causes delays in sound detection.

The findings also show that the time required for detecting the mobile equipment tends to be shorter than that of stationary equipment. (The exception is pile drivers and hammers because of the uniqueness of the sound they produce.) There are cases when mobile equipment is detected as stationary equipment at first before being recognized as a piece of mobile equipment. This can be attributed to the device latency and the similarity in acoustical patterns exhibited by different types of construction equipment. To reduce latency, the models need to be trained on small data frames. However, the use of small frames may cause an adverse effect on the model accuracy if the acoustical features important to a certain sound type cannot be found in one data frame, making it difficult for the sound classifier to distinguish between classes (Messi et al. 2020). Therefore, when the feature pattern has less variation among the sound classes, it is necessary to increase the frame length. We plan to train additional models with longer frame lengths.

Table 3. Result the experiments in the laboratory

Equipment Type	Probability	Duration (sec)
Forklift	0.903	5.222
Compactor roller	0.882	5.111
Grader	0.826	5.222
Hammer	0.783	5.111
Bulldozer	0.768	5.667
Scraper	0.681	5.444
Water truck	0.672	8.278
Excavator	0.656	4.000
Front end loader	0.627	7.333
Steel welding	0.604	9.000
Pile driver	0.601	3.111
Concrete pumper	0.532	9.667
Pneumatic breaker	0.486	7.778
Saw	0.454	7.889
Pneumatic tamper	0.413	7.667

The results from the experiments on the construction site (see Table 4) generally showed better performance--in both terms of probability and time to the first alert--when the worker was closer to the equipment. The probability range among the equipment is 27%-94% and 39%-93% when the worker is 10m and 20m, respectively, away from the equipment. Acoustic sensing in the field experiments was affected by surrounding equipment noise and human activities on the construction site. This could be the main cause of low probability for some types of equipment such as the excavator (27% at 10m) and the saw (51% at 10m). Regarding the latency, the longest duration was 14 seconds when the device detects the sound of the front-end loader, which is still relatively short. Compared with the experiments in the controlled environment, performance in the field tests was significantly reduced. The controlled environment is free of exterior noise, allowing easier and faster picking up of sound features and a higher confidence result.

Table 4. Average Result from Site Investigation

Equipment Type	Probability		Duration (sec)	
Distance	10m	20m	10m	20m
Excavator	0.27	0.39	7.67	11.33
Frontend loader	0.92	0.53	3.33	14.00
Hammer	0.94	0.93	7.67	10.00
Saw	0.51	0.66	9.00	10.67

Discussions

Heavy mobile equipment is a main source of fatal incidents in construction. Therefore, detecting the sound of mobile construction equipment is crucial to improving job site safety. Despite the significance of this issue, there is no system available that can detect auditory signals from vehicles that may create struck-by hazards, particularly when those sounds are buried in background noise. Early detection of the sound of mobile equipment can allow timely alerts to workers.

This study developed an AI model using CNN-based signal processing to enable early detection of auditory signals related to potential collision hazards. We trained and tested various models with different signal-to-noise ratios. To reduce the number of false negatives (missing any mobile equipment that will likely cause harm to field workers), the study emphasized the recall evaluation metric, which measures how many observations the model correctly predicted over the total number of observations. The findings (see Table 2)

showed that the model has a low probability of missing a potential hazard. Also, the precision value, which indicates the mislabeling of normal background sound as a danger, is in an acceptable range.

The efficacy of the proposed framework was evaluated with controlled and field experiments that aimed at assessing the suitability of the mobile application in reducing incidents on construction sites. The processing capacity of a smartphone obviously plays a great role in determining how fast it detects and assesses the hazards, and the quality of the built-in microphone is critical to the input sound. Moreover, the level of background noise greatly affects the efficacy of the framework. The results of the implementation experiments indicated that the probability of true positives for the controlled experiment is much greater than those of field tests, because the test sounds in the field tests were buried in background noises (e.g., near-by operations). Due to the scope of this Small Study, a comparison of our frameworks with existing electric systems for noise filtering (e.g., Etymotic Research Music-Pro Electronic Earplugs, Elvex COM-655 earplugs, 3M Peltor LEP-200, and Howard Leight Impact(R) Sport) was not completed but is in our future research plan. It is worth clarifying that the existing systems do not automatically distinguish between "important" and "bad" noises but filter sounds based on a predefined threshold of sound frequencies. Unlike the existing systems, our proposed framework is capable of automatically distinguishing between noises from mobile equipment and stationary equipment. Still, future experimental comparisons are needed to verify the efficacy and effectiveness of the new framework compared to other commercial devices.

While the time required to capture important signals tends to be longer in the field experiments, it is still in an acceptable range. In addition, the device still can quickly detect sounds with unique characteristics (like hammers) despite the presence of loud background noise. Lastly, weird sounds generated by equipment when it is not properly maintained may affect the sound and increase the possibility of making a wrong prediction. Training the models using data with more background noise may help improve the device's performance. The study demonstrated that CNN has a great potential for detecting important auditory signals buried in loud background noise on construction sites. However, the CNN model is considered a "black box" in which the learned equation difficult to explain to the end users because the model is comprised of complex relationships between numerous input features and the final output results from a large amount of data. This contrasts with traditional statistic models (e.g., linear regression and decision trees), which rely on a simple equation of a few variables that is easier to explain and interpret. The use of AI-equipped mobile devices requires high trust from the users because even though the model can approximate any functions represented by the data, studying its structure will not give any insights on the structure of the function being approximated. Machine learning models do not provide an explicit estimate of the importance of each feature on the model predictions. Also, is it difficult for the users to understand how different features interact.

Although this study proves that the proposed CNN model is a reliable technique to help detect potential collision hazards at the construction site, there are still areas to be improved for successful practical implementation. One particular limitation of this research is that the system could not capture the location of mobile equipment, and sound localization would help workers be aware of their position relative to the direction and distance to the hazard. Mobile vehicles moving toward a worker is a risk, for example, but not if moving away. Thus, localization is important to reduce false alarms for the system. In addition, the background noise considered in this study is limited to sounds from stationary equipment. Other types of background noises--such as natural sounds (e.g., wind, rain) and transportation (e.g., car engine, horn)--should be added to the dataset.

This study has provided a strong foundation on which to build more realistic models to detect collision hazards under the complicated nature of construction noises. Construction sites are dynamic environments where many activities are performed concurrently; site sound acoustic sensing is highly susceptible to noise and influences the properties of sound detected. Therefore, innovative approaches to de-noise and enhance the sound signals should be applied before the input sounds are being processed by our model in future work.

Changes/problems that resulted in deviation from the methods

N/A

Future funding plans

We plan to leverage the results from this small study to develop large-scale proposals to secure funding from national agencies, including the National Science Foundation (NSF) and the National Institute of Occupational Safety and Health (NIOSH).

Presentations, Publications, and Dissemination Plan

The findings from this study have been disseminated through a few peer-reviewed technical papers, including:

1. Huang*, Y., Trinh, M. T., & Le, T. (2021). Critical Factors Affecting Intention of Use of Augmented Hearing Protection Technology in Construction. ASCE Journal of Construction Engineering and Management, 147(8), 04021088.
2. Dang, K., Le, T. A Novel Audio-Based Machine Learning Model for Automated Detection of Collision Hazards at Construction Sites. 2020 ISARC conference, Japan (2020).
3. Kehinde Elelu , Tuyen Le (under review). Auditory Surveillance of Collision Hazards at the Construction Site Using Convolutional Neural Network. Automation in Construction.

References

J. W. Hinze and J. Teizer, "Visibility-related fatalities related to construction equipment," Saf. Sci., vol. 49, no. 5, pp. 709–718, Jun. 2011, doi: 10.1016/j.ssci.2011.01.007.

E. D. Marks and J. Teizer, "Method for testing proximity detection and alert technology for safe construction equipment operation," Constr. Manag. Econ., vol. 31, no. 6, pp. 636–646, Jun. 2013, doi: 10.1080/01446193.2013.783705.

T. C. Morata, C. L. Themann, R. F. Randolph, B. L. Verbsky, D. C. Byrne, and E. R. Reeves, "Working in noise with a hearing loss: perceptions from workers, supervisors, and hearing conservation program managers," Ear Hear., vol. 26, no. 6, pp. 529–545, Dec. 2005, doi: 10.1097/01.aud.0000188148.97046.b8.

E. Vinnik, P. M. Itskov, and E. Balaban, "Individual differences in sound-in-noise perception are related to the strength of short-latency neural responses to noise," PLoS One, vol. 6, no. 2, pp. 1–8, 2011, doi:10.1371/journal.pone.0017266.

C. F. Cheng, A. Rashidi, M. A. Davenport, and D. V. Anderson, "Activity analysis of construction equipment using audio signals and support vector machines," Autom. Constr., vol. 81, pp. 240–253, Sep. 2017, doi: 10.1016/j.autcon.2017.06.005.

C. F. Cheng, A. Rashidi, M. A. Davenport, and D. Anderson, "Audio signal processing for activity recognition of construction heavy equipment," in ISARC 2016 - 33rd International Symposium on Automation and Robotics in Construction, 2016, pp. 642–650, doi: 10.22260/isarc2016/0078.

Y. C. Lee, M. Shariatfar, A. Rashidi, and H. W. Lee, "Evidence-driven sound detection for prenotification and identification of construction safety hazards and accidents," Autom. Constr., vol. 113, p. 103127, May 2020, doi: 10.1016/j.autcon.2020.103127.

Y. Xie et al., "Historical Accident and Injury Database-Driven Audio-Based Autonomous Construction Safety Surveillance."

A. Perlman, R. Sacks, and R. Barak, “Hazard recognition and risk perception in construction,” Saf. Sci., vol. 64, pp. 13–21, Apr. 2014, doi: 10.1016/j.ssci.2013.11.019.

C. A. Sabillon, A. Rashidi, B. Samanta, C. F. Cheng, M. A. Davenport, and D. V. Anderson, “A productivity forecasting system for construction cyclic operations using audio signals and a Bayesian approach,” in Construction Research Congress 2018: Construction Information Technology - Selected Papers from the Construction Research Congress 2018, 2018, vol. 2018-April, pp. 295–304, doi: 10.1061/9780784481264.029.

Sherafat, Rashidi, Lee, and Ahn, “A Hybrid Kinematic-Acoustic System for Automated Activity Detection of Construction Equipment,” Sensors, vol. 19, no. 19, p. 4286, Oct. 2019, doi: 10.3390/s19194286.

T. Zhang, Y. C. Lee, M. Scarpiniti, and A. Uncini, “A supervised machine learning-based sound identification for construction activity monitoring and performance evaluation,” in Construction Research Congress 2018: Construction Information Technology - Selected Papers from the Construction Research Congress 2018, 2018, vol. 2018-April, pp. 358–366, doi: 10.1061/9780784481264.035.

Y. Lee, D. K. Han, and H. Ko, “Acoustic signal based abnormal event detection in indoor environment using multiclass adaboost,” IEEE Trans. Consum. Electron., vol. 59, no. 3, pp. 615–622, 2013, doi: 10.1109/TCE.2013.6626247.

M. Cowling and R. Sitte, “Comparison of techniques for environmental sound recognition,” Pattern Recognit. Lett., vol. 24, no. 15, pp. 2895–2907, Nov. 2003, doi: 10.1016/S0167-8655(03)00147-8.

A. J. Eronen et al., “Audio-based context recognition,” in IEEE Transactions on Audio, Speech and Language Processing, 2006, vol. 14, no. 1, pp. 321–329, doi: 10.1109/TSA.2005.854103.

A. Wieczorkowska, E. Kubera, T. Słowik, and K. Skrzypiec, “Spectral features for audio based vehicle and engine classification,” J. Intell. Inf. Syst., vol. 50, no. 2, pp. 265–290, Apr. 2018, doi: 10.1007/s10844-017-0459-2.

N. Alamdari, F. Saki, A. Sehgal, and N. Kehtarnavaz, “An unsupervised noise classification smartphone app for hearing improvement devices,” in Signal Processing in Medicine and Biology Symposium (SPMB), 2017 IEEE, 2017, vol. 2018-Janua, pp. 1–5, doi: 10.1109/SPMB.2017.8257031.

M. Wirz, D. Roggen, and G. Tröster, “A wearable, ambient sound-based approach for infrastructureless fuzzy proximity estimation,” in Proceedings - International Symposium on Wearable Computers, ISWC, 2010, doi: 10.1109/ISWC.2010.5665863.

S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, “Indoor localization without infrastructure using the acoustic background spectrum,” in MobiSys’11 - Compilation Proceedings of the 9th International Conference on Mobile Systems, Applications and Services and Co-located Workshops, 2011, doi: 10.1145/1999995.2000047.

Z. H. Janjua, M. Vecchio, M. Antonini, and F. Antonelli, “IRESE: An intelligent rare-event detection system using unsupervised learning on the IoT edge,” Eng. Appl. Artif. Intell., vol. 84, pp. 41–50, Sep. 2019, doi: 10.1016/j.engappai.2019.05.011.

Samantha Brown, William Harris, Raina D. Brooks, Xiuwen Sue Dong. “Fatal Injury Trends in

the Construction Industry." CPWR–The Center for Construction Research and Training, Page 3, 2021. https://www.cpwr.com/wp-content/uploads/DataBulletin-February-2021.pdf

Singh, N., Khan, R.A. and Shree, R., 2012. MFCC and prosodic feature extraction techniques: a comparative study. International Journal of Computer Applications, 54(1). Link https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.258.9750&rep=rep1&type=pdf

Roslidar, R., Saddami, K., Arnia, F., Syukri, M. and Munadi, K., 2019, August. A study of fine-tuning CNN models based on thermal imaging for breast cancer classification. In 2019 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom) (pp. 77-81). IEEE. Link https://ieeexplore.ieee.org/document/8875661

Messi, L., Naticchia, B., Carbonari, A., Ridolfi, L. and Di Giuda, G.M., 2020, July. Development of a Digital Twin Model for Real-Time Assessment of Collision Hazards. In Creative Construction e-Conference 2020 (pp. 14-19). Budapest University of Technology and Economics. Link https://www.iaarc.org/publications/fulltext/ISARC_2020_Paper_201.pdf

Huang, Yongcan, Minh Tri Trinh, and Tuyen Le. "Critical Factors Affecting Intention of Use of Augmented Hearing Protection Technology in Construction." Journal of Construction Engineering and Management 147, no. 8 (2021): 04021088. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002116

Dang, Khang, and Tuyen Le. "A Novel Audio-Based Machine Learning Model for Automated Detection of Collision Hazards at Construction Sites." In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 37, pp. 829-835. IAARC Publications, 2020. DOI: https://doi.org/10.22260/ISARC2020/0114

Wessel, David, and Matthew Wright. "2001: Problems and Prospects for Intimate Musical Control of Computers." In A NIME Reader, pp. 15-27. Springer, Cham, 2017. Link https://cnmat.berkeley.edu/sites/default/files/attachments/2002_problems-and-prospects-for-intimate-musical-control-of-computers.pdf

CPWR THE CENTER FOR CONSTRUCTION RESEARCH AND TRAINING

www.cpwr.com • www.elcosh.org