


Article

A Hybrid Kinematic-Acoustic System for Automated Activity Detection of Construction Equipment

Behnam Sherafat ^{1,*}, Abbas Rashidi ¹, Yong-Cheol Lee ² and Changbum R. Ahn ³

¹ Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, UT 84112, USA; abbas.rashidi@utah.edu

² Department of Construction Management, Louisiana State University, Baton Rouge, LA 70803, USA; yclee@lsu.edu

³ Department of Construction Science, Texas A&M University, College Station, TX 77843, USA; ryanahn@tamu.edu

* Correspondence: behnam.sherafat@utah.edu; Tel.: +1-801-558-1001

Received: 13 August 2019; Accepted: 1 October 2019; Published: 3 October 2019



Abstract: Automatically recognizing and tracking construction equipment activities is the first step towards performance monitoring of a job site. Recognizing equipment activities helps construction managers to detect the equipment downtime/idle time in a real-time framework, estimate the productivity rate of each equipment based on its progress, and efficiently evaluate the cycle time of each activity. Thus, it leads to project cost reduction and time schedule improvement. Previous studies on this topic have been based on single sources of data (e.g., kinematic, audio, video signals) for automated activity-detection purposes. However, relying on only one source of data is not appropriate, as the selected data source may not be applicable under certain conditions and fails to provide accurate results. To tackle this issue, the authors propose a hybrid system for recognizing multiple activities of construction equipment. The system integrates two major sources of data—audio and kinematic—through implementing a robust data fusion procedure. The presented system includes recording audio and kinematic signals, preprocessing data, extracting several features, as well as dimension reduction, feature fusion, equipment activity classification using Support Vector Machines (SVM), and smoothing labels. The proposed system was implemented in several case studies (i.e., ten different types and equipment models operating at various construction job sites) and the results indicate that a hybrid system is capable of providing up to 20% more accurate results, compared to cases using individual sources of data.

Keywords: construction equipment; audio and kinematic signals; sensor fusion; dimension reduction; activity detection; support vector machines

1. Introduction

In the construction industry, three important resources at job sites are crews, materials, and equipment. For the majority of construction projects, a major expense item is the budget allocated to the acquisition, rental, and maintenance of heavy equipment [1]. The costs depend on several factors, including the idle and working hours, as well as productivity and efficiency rates. When it comes to working hours and productivity rates, two types of problems can occur during equipment working shifts. One is idle time, or the lack of operation while the equipment engine is working. This situation can result from poor planning (e.g., lack of a sufficient number of dump trucks to be loaded by an excavator), human factors, etc. The second type of problem is downtime when equipment is inactive, due to mechanical/electrical problems and the need to be repaired.

Carefully analyzing and monitoring heavy equipment productivity rates and monitoring downtimes, idle times, and productive times are well-known significant factors in the success of construction projects. Traditional methods for productivity analysis and construction equipment and machinery performance monitoring are through direct observations, as well as conducting interviews and surveys. These methods are labor-intensive and prone to error, making them impractical for larger job sites and more complex environments. As a result, there is an increasing demand for more efficient and systematic solutions for productivity analysis of heavy equipment under various conditions [2].

Construction equipment productivity rates are directly associated with various activities performed by the machine during routine operations. Recognizing these activities is the first step toward analyzing efficiency rates [3,4]. A systematic activity recognition system could help project managers (1) evaluate machines idle versus working times, (2) estimate cycle times of construction operations, and (3) analyze productivity rates [5,6].

Recent advancements in technology motivated researchers to develop automated systems to decrease human interference [7–9]. In parallel with recent advancements in developing computing methods and Information Technology (IT) tools, researchers have devised several techniques for automated equipment activity detection in construction job sites. These techniques could be divided into three major categories. The first category uses vision-based technologies, such as two-dimensional (2D)/image and three-dimensional (3D) range cameras, and utilizes computer vision algorithms to analyze images and videos [10–12]. The second category employs kinematic sensors, such as accelerometers and gyroscopes, to record the equipment's kinematic signals and detect its activities [13,14]. The third method is based on processing audio signals [1,15,16], using single microphones or microphone arrays to record sound patterns generated by machines while performing various operations.

The studies mentioned above use one of the three major sources of data (i.e., visual, kinematic, or acoustic data) to detect equipment activities. Each type of data, as well as the associated processing algorithms, have certain limitations and drawbacks. For computer vision-based methods, factors such as digital cameras' limited field of view, illuminations, and occlusions, could negatively impact the widespread use of this category at complex construction job sites. For the case of audio-based methods, certain types of construction machinery do not generate distinct sound patterns during the course of operation (e.g., cranes, soil compactors, etc.), so relying on audio as the only source of data might not always be a feasible solution. Finally, for the kinematic-based methods, it could be challenging to directly attach kinematic sensors to the body of the machine (e.g., jackhammers, concrete pumps, concrete truck mixers, etc.).

To address the distinct weaknesses of these existing methods, the authors have investigated the feasibility of integrating two major sources of data, kinematic and acoustic, and developing a hybrid system for activity detection of construction equipment. The proposed hybrid system is based on a combined feature-based approach to increase the final results' accuracy. Several types of features are extracted from both data sources and merged to generate more accurate results. The presented system is capable of detecting a wide range of activities taking place at construction job sites, such as moving forward/backward, arm raising/lowering, rotating, loading and excavating, etc. Compared to using individual sources of data, a hybrid system can cover a broader range of equipment types and job site conditions, eventually generating more accurate results.

This paper is organized as follows: Section 2 presents a broad overview of activity detection methods in different areas. A more in-depth literature review on various methods for construction equipment activity detection will also appear in this section. In Section 3, we provide technical details about the proposed hybrid system. Details of experimental setups and obtained results will be presented in Section 4. Finally, Section 5 includes the conclusions and lessons learned from this study.

2. Literature Review

As indicated earlier, two major methods for equipment activity detection are based on kinematic or audio signals. In the following sections, various studies in the general field of audio recognition

and activity detection are examined. Then, we discuss recent studies on equipment activity detection using audio data. In the following section, we investigate general activity detection studies using kinematic signals. These studies have utilized various types of features, machine learning models, and evaluation processes. We also review more related studies in the domain of construction, especially equipment activity detection. Finally, in the last section, we provide a comprehensive comparison between these two methods and discuss the role of data fusion.

2.1. Audio Recognition and Activity Detection Using Audio Signals

Audio carries useful information about our surroundings. This particular source of data could be used to identify events that occur in our environment for Enhanced Living Environments (ELE) and Ambient Assisted Living (AAL) [17]. With the advent of new methods and tools, sound recognition and classification have attracted a significant amount of attention from researchers and practitioners. Audio recognition and classification consist of extracting different features from a sample audio file and feeding these features into a machine-learning algorithm, to detect classes of the present sounds. This topic has been studied for ambient sound classification [18–20], noise signal classification [21,22], speech/music classification [23], music genre classification [24], human accent or language classification, speaker recognition [25,26], and indoor localization [27]. Hinton et al. [28] and McLoughlin et al. [29] used Deep Neural Network (DNN) to develop an automated speech recognition system and a robust sound event classification, respectively. Similarly, Graves et al. [30] utilized Deep Recurrent Neural Networks (DRNN) for speech recognition and sound event classification. Hwang and Lee [31] developed a crowdsourcing framework to obtain real-life data from the environment to detect different types of activities occurring in the surroundings. The k-nearest neighbors' method has been used in their system, however, their reported accuracies are not sufficiently high.

Few studies have been conducted on equipment activity detection using audio signals. In References [1,2,15,16,32,33], audio signals have been used to detect construction equipment activities. In those studies, major activities were defined as those directly contributing to the productivity of the project (e.g., loading, pushing soil, excavating, etc.). Also, those studies only utilize the magnitude of Short Time Fourier Transform (STFT) of audio signals as the input feature, and the proposed systems are only applicable for equipment that generates distinctive sound patterns. This situation is additionally problematic since the presence of any sound barriers might affect the output's accuracy. In a complementary effort to fix some of those issues, Cheng et al. [2] studied necessary hardware and software settings for an audio-based equipment activity detection system to obtain more rigorous results. They compared different types of microphones and locations to achieve optimal solutions for their system. Cheng et al. [34] proposed an approach to label weakly labeled data, based on the clustering behavior of audio features in data frames.

2.2. Activity Detection Using Kinematic Signals

Unlike audio, kinematic data has been commonly utilized for activity recognition purposes in different areas. Kinematic sensors are cost-effective and easy-to-use devices which can be conveniently used for activity detection purposes, especially detecting human beings' various activities. Several studies have been conducted in the last two decades regarding human activity detection [35–42]. Since human activity detection is not the focus of this paper, only a few recent papers on this subject are discussed in this section. In one of the most relevant studies, Ignatov and Strijov [43] utilized smartphone accelerometer data and the fundamental periods' extraction method to detect six types of human activities. Another application of human activity detection is gait analysis. Anwary et al. [44] created an app for this purpose after investigating different locations for optimal sensor placement. They used both accelerometer and gyroscope data and extracted ten different features. Their results indicated that the orientation and location of sensors could significantly impact the results. Lee and Kwan [45] integrated Global Positioning System (GPS) and accelerometer data to detect the location

and type of human activities. They extracted 59 features and applied two different approaches, i.e., 10-fold cross-validation and random forest.

In recent years, several studies have been conducted to detect equipment activities and measure the performance of construction operations using kinematic sensors. Ahn et al. [46] proposed a monitoring system to determine the equipment's operational efficiency and environmental performance based on kinematic signals. By observing signal energy, idle periods of equipment are determined. This research only uses signal energy as a feature and does not use any machine learning models or signal processing techniques. Moreover, it was tested on one piece of equipment (e.g., excavator), and cannot be generalized to all other types of equipment in real-world construction job sites. Akhavian and Behzadan [47] developed an automated model to track equipment and determine its different motions. The model detects relative movements of different parts of the equipment using magnetic field and tilt sensing and creates a real-time 3D animation. The output of their system can be used for long-term planning and does not detect various activities handled by the equipment for performance monitoring. Later, in References [48,49], Akhavian and Behzadan fused different data types—such as weight, position, and orientation—to detect varying equipment activities' duration and also determine properties of the queuing system to be used in a simulation model. Through a separate research study [46], the authors utilized accelerometer, gyroscope, and GPS data for detecting heavy construction equipment activities. They have not reported any pre-processing methods on sensor data and used a few time-domain and frequency-domain features, which led to moderate accuracies that may require further improvements. In a similar effort, Ahn et al. [14] developed a method based on accelerometer data to detect three different classes of activities: engine-off, idling, and working modes. Their proposed method was preliminarily tested on excavators. In one of the most recent studies, Kim et al. [50] implemented Inertial Measurement Units (IMUs) to detect an excavator's various activities. They used the concept of Dynamic Time Wrapping (DTW) to increase the accuracy rates to detect mixed activities. This research mainly focused on cabin rotation to determine process cycle time.

2.3. Audio and Kinematic Data Comparison

As indicated earlier, the existing studies in the literature often used either audio or kinematic data as the primary source of information for automated activity detection of construction equipment. There are some limitations and restrictions in using each of these data sources. Due to sound barriers and long distances between the microphone and the sound's source, audio might not be a useful source in large job sites. Moreover, some newer equipment models perform quietly and might not generate distinctive sound patterns necessary for further processing. A similar issue may exist for kinematic signals where new equipment might not generate strong vibration signals. Also, some types of equipment do not offer proper space to place kinematic sensors or mobile phones (e.g., small hand drills). Another issue within the existing methods is that in most current studies, only a few feature types are selected and extracted. Many of them have used STFT coefficients, which might not be sufficient to achieve accurate results. Evaluating different types of features, and selecting the most useful ones, is one of the advantages of this paper's proposed activity detection framework.

The concept of data fusion has previously been used by researchers to integrate different types of data and overcome the limitations of merely using single data sources. Garcia-Ceja et al. [51] fused audio and sensor data using a multi-view stacking method to recognize different types of human activities. They suggested that the multi-view stacking method generates better results compared to aggregating features. They have used their method for human activity detection, which occurs in less noisy and complex environments as compared to construction job sites.

The presented research study in this paper is the first attempt towards using audio-kinematic sensor fusion to automatically detect multiple equipment activities with high accuracy. The contribution of this work to the existing body of knowledge is two-fold: first, by fusing audio with kinematic signals for multiple activity detection, and second, by using different types of time-domain and frequency-domain features.

3. Materials and Methods

The proposed hybrid system in this paper utilizes audio and kinematic data to detect multiple activities of construction equipment. Audio recordings from a microphone (placed outside the cabin) and kinematic recordings from IMUs (accelerometer and gyroscope embedded in a mobile phone placed inside the cabin) are used together to obtain accurate results. As explained in the previous section, each of the audio and kinematic signals has specific advantages and drawbacks. Microphones can cover large areas of job sites without needing to be directly attached to a specific piece of machinery. It is also well known that audio patterns generated by construction equipment are often independent of the operator and the orientations and directions followed by the machine [1]. On the other hand, IMU sensors shall be connected directly to the equipment and are capable of detecting almost every minor motion of the machine during operation. As a result, integrating these two data sources will help overcome the drawbacks of each one while taking advantage of the other source of data.

As indicated in Figure 1, the proposed hybrid system consists of the following major components: (1) recording both audio and kinematic signals, (2) pre-processing data using de-noising algorithms, (3) selecting and extracting different audio- and kinematic-based features, (4) filtering less-useful features or reducing the dimension of the extracted feature sets, (5) sensor fusion using both audio and kinematic data, (6) training and testing a Support Vector Machines (SVM) model using extracted features, and (7) smoothing labels. These steps are explained in more detail in subsequent sections.

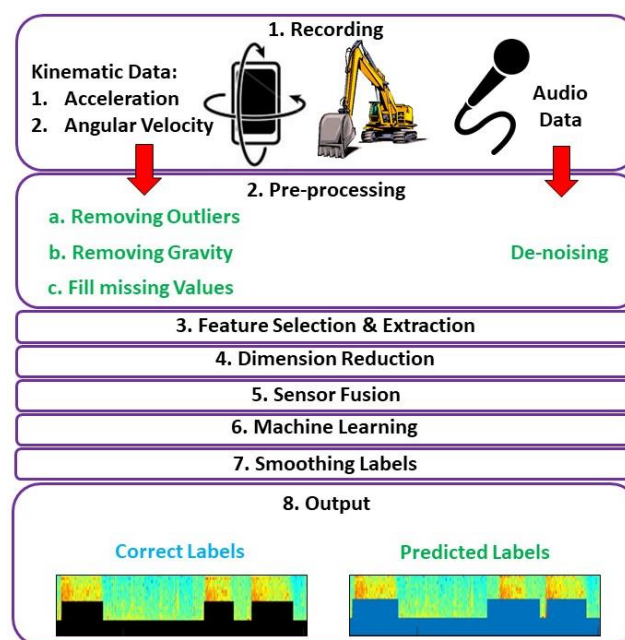


Figure 1. An overview of the hybrid kinematic-acoustic system for activity detection of construction machinery.

3.1. Recording Audio and Kinematic Signals

The first step toward implementing the presented activity detection system is collecting and recording input data. In this paper, the authors have recorded two types of data: (1) audio: This data is recorded using a microphone placed outside the cabin, and (2) kinematic: This data is recorded using IMUs embedded in a mobile phone inside the cabin. Technical details about the devices used in this paper are described in the experimental setup section. In this paper, two types of kinematic data are used: (1) acceleration data (A_x , A_y , and A_z), and (2) angular velocity data (V_x , V_y , and V_z). The following illustration shows the X, Y, and Z orientation axes relative to a typical mobile phone (Figure 2). Kinematic data is recorded using accelerometer and gyroscope sensors embedded in the mobile device. Because the starting point of audio and kinematic data might be slightly different for time synchronization between audio and kinematic signals, an equipment horn has been used to

generate a sound and vibration. This short high-pitched sound is easily recognizable throughout the signal and could be used as the starting point of recording for synchronization. Also, a video camera constantly records the entire scene, capturing all of the activities performed by the equipment over time. This video will be further used as the benchmark in order to label actual equipment activities.

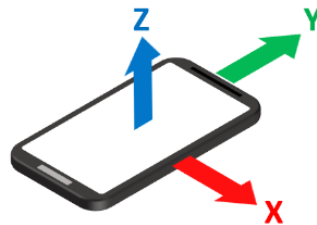


Figure 2. X, Y, and Z orientation axes relative to a typical mobile phone.

3.2. Pre-Processing Data

The recorded data contains noisy and unintended signals (e.g., gravity for kinematic data) which could negatively affect the proposed activity detection system's performance. These unwanted signals can be eliminated or diminished by utilizing some efficient algorithms. In this research, two different algorithms are used separately for the audio and kinematic signals.

To reduce unwanted background noise from the audio signal, the authors have implemented a signal enhancement algorithm developed by Rangachari and Loizou [52]. This algorithm has been used in similar studies [1] and it is proven to be an efficient method. Their method is capable of reducing noises in complex environments along with minimum distortion of the desired signals.

A similar process is also required for kinematic signals. The developed noise reduction process for kinematic signals consists of the following steps:

1. **Removing gravity:** This step is only applied on an accelerometer sensor because its values are subject to both dynamic (or external) and static (or gravity) accelerations. Thus, gravity components need to be eliminated from the signal. Equation (1) shows a low-pass filter for accelerometer sensor values. The cut-off frequency of 0.1 to 0.5 Hz is recommended to remove the gravity component from the data [53,54]. This equation calculates g-values for the sensor (g is initially set to zero) and then in Equation (2), g-values are subtracted from the sensor values. More details about these equations can be found in Bayat et al. [55].

$$g(t) = (1 - a) \times g(t - 1) + a \times s(t) \text{ and } g(0) = 0, \quad (1)$$

$$s(t) = s(t) - g(t), \quad (2)$$

where, $g(t)$ is the gravity value, which is initially set to zero and is updated accordingly, the parameter s is accelerometer sensor amplitude, the parameter t is time, and the parameter a is a variable between 0 and 1 that controls the cutoff of the filter. In similar applications for removing gravity, the value of a is set to 0.1. By implementing these equations, the effect of gravity is removed and the actual kinematic values representing relative movements with respect to the equipment's cabin can be derived.

2. **Removing outliers:** Outliers are unwanted noise or behaviors significantly different from the rest of the data. They decrease the accuracy of the system, so they should be eliminated. Data smoothing consists of techniques for removing these data points. Moving window methods are utilized to analyze data in smaller groups at a time. In this paper, the authors tested the moving window medians of length 3, 6, 12, and 24 using cross-validation. It is found that the window size of 3 is more effective to detect outliers.

3. Filling missing values: During recording data, and due to connection issues, some data points might be missed or not be recorded. These missing values affect the accuracy of the system. Thus, it is crucial to find a way to fill these values. The method for filling missing values is the same as detecting and removing outliers. Similar to the chosen window size for removing outliers, the authors tested different window sizes and found that window size 24 has less impact on the data.

Figure 3 demonstrates sample plots for acceleration (A_x) and audio signals with their respective spectrograms before and after de-noising for a jackhammer 305.5E2.

After both audio and kinematic signals are de-noised and refined, they need to be synchronized. In this paper, the authors have synchronized data manually. Using the generated sound and vibration from the equipment horn, a similar signal spike is detected on both audio and kinematic signals. Also, their respective spectrograms were identified and previous points were then cropped and shifted in a way that allowed both signals to have the same starting point. As a result, both signals were synchronized and could be further processed, as described in the following sections.

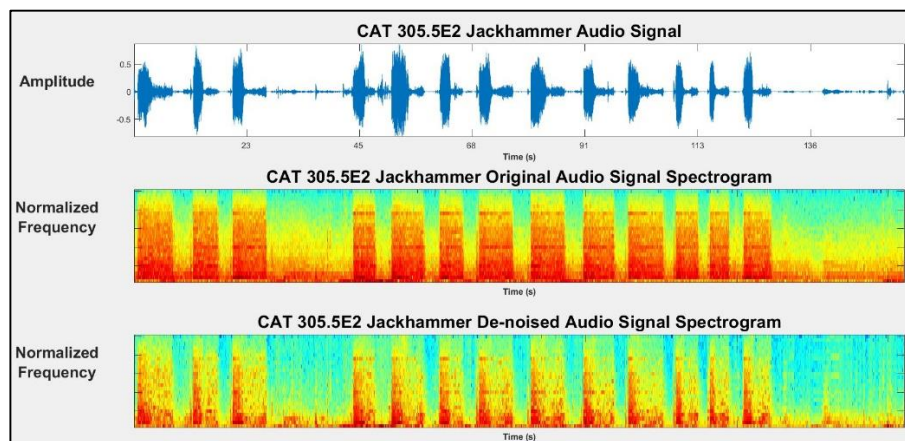


Figure 3. A sample of an audio signal and its respective spectrograms before and after de-noising (machine: jackhammer 305.5E2).

3.3. Selecting and Extracting Features

The same number of bins are required for feature fusion in the next steps to integrate both audio and kinematic signals. Because of different sampling frequencies of audio and kinematic signals (44,100 Hz for the audio signal and 100 Hz for the kinematic signal), the corresponding number of data points also differ. Thus, in order to obtain the same number of bins, two options are available: (1) the audio data can be down-sampled, or (2) the kinematic data can be up-sampled. These two options are evaluated and it is found that the up-sampled signal has the same amount of information [56], but down-sampling the audio data leads to information loss and signal distortion. All of the frequency components greater than half the new sampling rate ($\geq \frac{f_s^{New} (=100 \text{ Hz})}{2} = 50 \text{ Hz}$) need to be removed using a low-pass filter before applying the down-sampling process to avoid aliasing. This process removes most of the useful frequencies in the audio signal and distorts the signal. In Figure 4, a comparison between the down-sampling and up-sampling effect is shown. The original signal and the resampled signal for both scenarios are shown. It is clear that the down-sampled audio signal is distorted. Also, the authors evaluated both methods and found that down-sampling the audio signal decreases the accuracy of the model up to 20% with respect to up-sampling the kinematic signals. Therefore, the linear data interpolation function called “interp1” in MATLAB is used to up-sample the kinematic signal, from 100 Hz to 44,100 Hz.

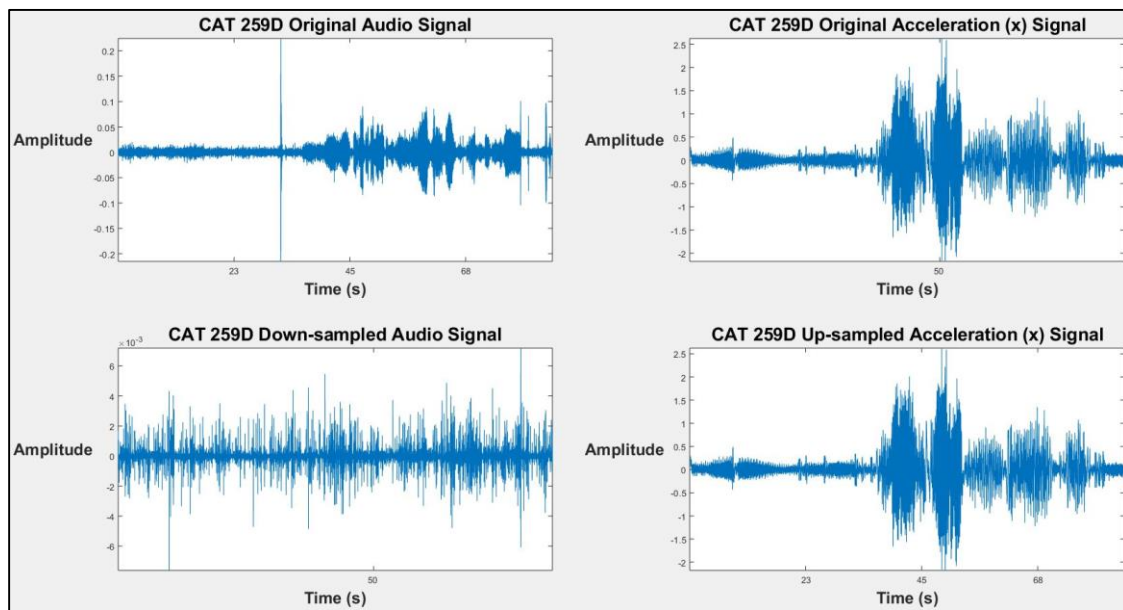


Figure 4. The original and resampled signals for both down-sampling and up-sampling options (machine: Caterpillar (CAT) 259D).

Interpolation is used to make the data points equivalent for both signals. The Linear 1-D data interpolation function called “interp1” in MATLAB is used to increase the number of data points for kinematic signal, from 100 data points per second to 44,100 data points per second. These new interpolated data points, which now both have a sampling frequency of 44,100, are used for further processes.

Now, both audio and kinematic signals have the same number of data points. For feature extraction, audio and kinematic signals were divided into short time segments and these segments were converted into a time-frequency domain representation using the STFT method. In this paper, STFT implementation in MATLAB using a Hanning window size of 512, with 50% overlap, was utilized to extract features, as described in previous papers [1,2,57]. This particular window size was chosen because it provides enough time resolution, as indicated by Cheng et al. [1]. Also, only the magnitudes of Short Time Fourier Transform (STFT) coefficients have been considered as features.

Another challenge during the feature extraction procedure is that some features work fine with one of the audio or kinematic data, while other features could work well with both. To take this point into account, different types of features in the time and frequency domains were considered for each of these signals, as shown in Table 1. These features were selected and tested for different types of equipment and job sites, and have been proven effective in class separability. STFT Coefficients have been used in Reference [1] and demonstrated satisfactory performance in the system. Root Mean Square (RMS) and Short Time Energy (STE) are related to signal energy, which had already shown accurate results in previous studies [58]. As described in Reference [59], the engine sound often has significant spectral components, which could help better identify different machine activities. Spectral Centroid (SC), Spectral Roll-Off (SRO), and Zero Crossing Rate (ZCR) were chosen in a previous research project for automatic vehicle and engine classification based on audio data [60]. Spectral Flux (SF) is another desired feature that has been used in similar research for vehicle engine classification [61]. All of these spectral features, such as Spectral Entropy (SE), were previously implemented to detect the presence of and identify specific types of vehicles in traffic using acoustic signals [62].

Table 1. Extracted features.

Audio and Kinematic	Only Kinematic
25 Short Time Fourier Transform (STFT) Coefficients	Zero Crossing Rate (ZCR)
Root Mean Square (RMS)	
Short Time Energy (STE)	
Spectral Flux (SF)	
Spectral Entropy (SE)	
Spectral Centroid (SC)	
Spectral Roll-Off (SRO)	

3.4. Dimension Reduction

In the Machine Learning area, dimension reduction is the process of reducing the number of features extracted from the training data. This process is particularly important for a robust activity detection and monitoring system, as the results should be eventually generated in real-time (or near real-time setting). To achieve this goal, and as demonstrated in Figure 5, the Principal Component Analysis (PCA) algorithm has been implemented to reduce the dimension of the training data. PCA is an efficient method commonly used to highlight variations and bring out strong patterns in a dataset, by transforming the feature space using centering and rotation such that the resulting vectors are pointing in the direction with the highest variance in descending order. Walse et al. [63] utilized PCA before feeding the features into a DNN model for human activity recognition using mobile sensors data. They have reported that PCA can decrease the computational time significantly, which can further be used for real-time purposes. Before using PCA, the extracted features are normalized using z-score to avoid scaling effects, while ensuring that any feature with larger domain will not dominate features with smaller domain [64–66]. A feature value X of a feature F is normalized to X' using Equation (3):

$$X' = \frac{X - \mu(F)}{\sigma(F)} \quad (3)$$

where, X is the feature value, $\mu(F)$ is the arithmetic mean of all values of feature F , $\sigma(F)$ is the standard deviation of all values of feature F , and X' is the normalized feature value.

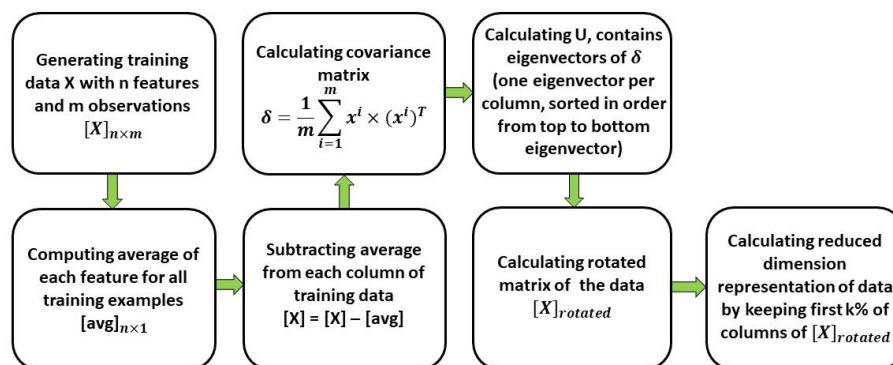


Figure 5. The proposed dimension reduction algorithm using Principal Component Analysis (PCA).

PCA aims at finding the most discriminate principal components in the orthonormal eigenvector space. The eigenvalues of the covariance matrix of the data provide the variance at each orthogonal direction (eigenvectors). In this paper, 32 features from each kinematic data type (totally 6×32) and 31 features from audio data have been extracted. As a rule of thumb, the Kaiser–Guttman rule [67] states that the principal components with eigenvalues greater than 1 should be retained. In this paper, the authors used this rule, and as a result, we found that 13 principal components have eigenvalues greater than 1, which cumulatively explain 95.12% of the variance. In Figure 6, the top plot shows the principal components and their eigenvalues. The eigenvalues for the first 13 components are greater than 1. Also, in the bottom plot, the principal components and the explained variance for each of them is shown. These components are used for the next steps.

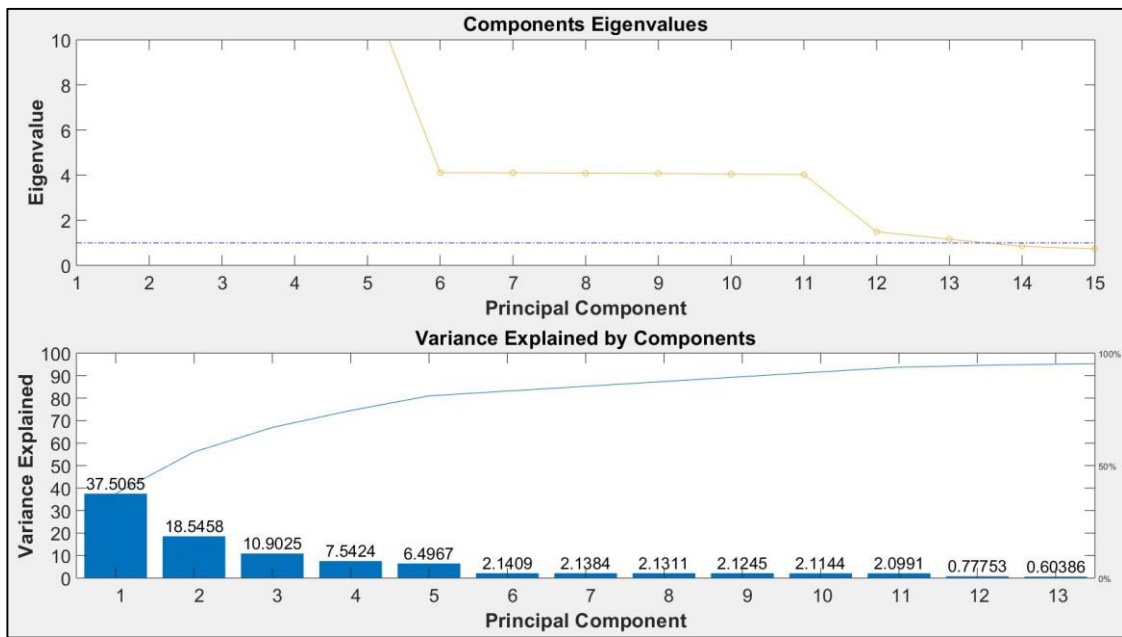


Figure 6. Most discriminative principal components and the variance they explain.

3.5. Sensor Fusion

In order to evaluate the performance of the proposed hybrid system, three scenarios were taken into account: (1) using audio signals only, (2) using kinematic signals only, and (3) integrating audio and kinematic signals. For the first and second scenarios, only audio and kinematic features were extracted respectively, and fed into the model. For the third scenario, a feature fusion procedure was implemented to integrate both audio and kinematic signals for training and testing the SVM model. In other words, extracted feature matrices from audio and kinematic data are combined to obtain a single feature matrix, which is more discriminative than any of the input feature vectors. It is a common practice to fuse multiple sources such as audio, video, and text to improve the accuracy of the system and generate useful information. The most common method for sensor fusion in the domain of wearable sensors is aggregation, which means that instead of training classification models for each data type, different feature sets from different sources are concatenated to obtain a single feature set for training a single classification model. Figure 7 demonstrates the process of sensor fusion using the feature aggregation method. In Figure 7, columns show the feature sets and rows representing the observations (data points). Parameters “m” and “g” show the number of data points for audio and kinematic data, which are equal using the method explained in Section 3.3. As explained in the next section, this feature set was further used to train and test the SVM model. A detailed explanation of this fusion process can be found in Reference [51].

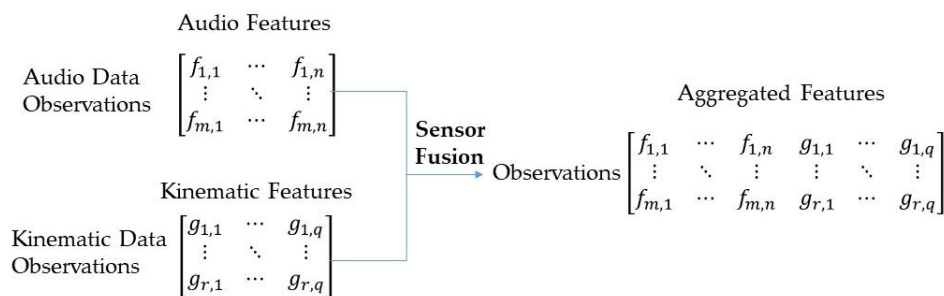


Figure 7. Sensor fusion using the feature aggregation method.

3.6. Support Vector Machines (SVM) Model

An SVM model with Radial Basis Function (RBF) kernel was utilized for predicting activities of construction machines. This machine learning classifier has been tested in previous studies conducted by the authors [1,2,57] and was found efficient and accurate. For comparison purposes and by utilizing the video recordings as a benchmark, all different periods of audio and kinematic signals were identified and labeled accordingly. Table 2 demonstrates a chronological list of activities performed by a CAT 259D compact track loader. A visual representation of the corresponding activities for this specific case study is also depicted in Figure 8. The black bars on the bottom plot of Figure 8 are the actual labels for activities 1 to 4. It is necessary to mention that the authors used 90% of available data for training and the remaining 10% for testing the developed machine learning classifier. Generated labels in this step are fed into the next step for post-processing.

Table 2. CAT 259D activities and the corresponding list of actual labels.

Activity	Start (s)	End (s)	Duration (s)	Activity Label
Arm Raising	0	13.4	13.4	2
Arm Lowering	13.4	24.4	11	2
Shovel Lowering	24.4	30.4	6	2
Shovel Raising	30.4	35.4	5	2
Arm Lowering	35.4	36.5	1.1	2
Moving Forward	36.5	45.97	9.47	3
Minor Stop	45.97	46.84	0.87	1
Moving Backward	46.84	52.6	5.76	3
Minor Stop	52.6	53.5	0.9	1
Turning Right	53.5	60.85	7.35	4
Minor Stop	60.85	61.65	0.8	1
Turning Left	61.65	70.13	8.48	4
Arm Raising	70.13	72.89	2.76	2
Turning Right	72.89	77.16	4.27	4
Arm Lowering	77.16	78.84	1.68	2
Shovel Lowering	78.84	80.47	1.63	2
Stop	80.47	87	6.53	1
End	87			

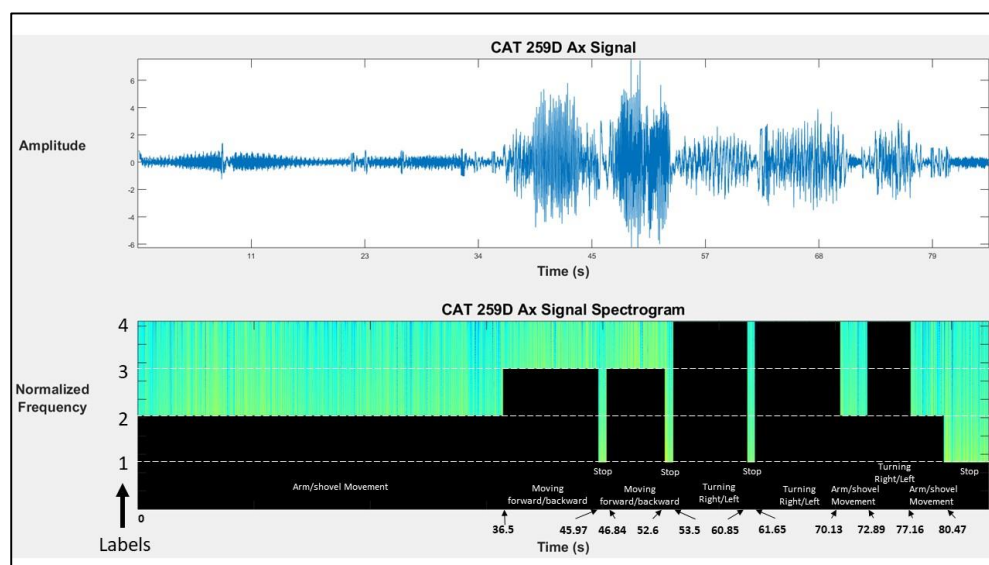


Figure 8. CAT 259D actual labels.

3.7. Smoothing Labels

Because the labels generated by SVM are fluctuating, they cannot precisely represent their related activities. So, in order to visualize the predicted labels accurately, three smoothing methods were applied: (1) Small Window Filtering (SWF), (2) Big Window Filtering (BWF), and (3) Markov Chain Filtering (MCF). The y -axis shows the labels corresponding to their related activities and the x -axis shows the time bins. The first two algorithms use Moving Average Window (MAW) and the only difference between them is the size of the window utilized by each of them. The window sizes for SWF and BWF are determined as 2 and 6 with a threshold of 0.68. These window sizes and threshold are chosen based on 10-fold cross-validation, which leads to the most accurate results. As a simple example, the process of window filtering is shown in Figure 9. In this figure, the window size of 6 is shown. The algorithm calculates the ratio in Equation (4), which is $\frac{5}{6} = 0.83$, and changes its label from “Label 1” to “Label 2”, because it has a ratio higher than 0.68.

$$\text{Smoothed Label} = \begin{cases} \text{Label 2,} & \text{if ratio} = \frac{\text{Count of Label 2 in the Window}}{\text{Total Count of Labels in the Window}} \geq 0.68 \\ \text{Label 1,} & \text{otherwise} \end{cases} \quad (4)$$

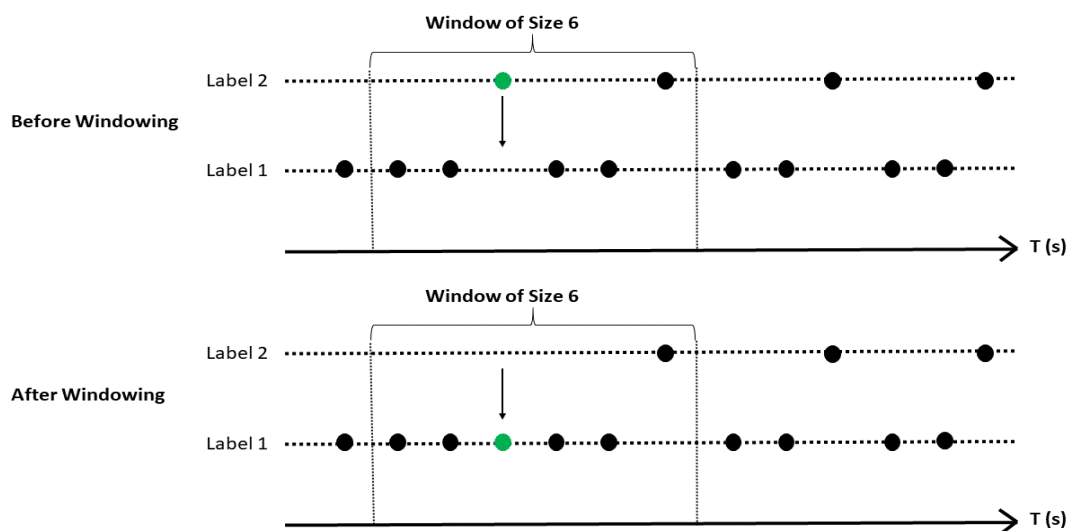


Figure 9. A sample window filtering process.

The third method uses the MCF to predict the next label, based on the previous label state. MCF scans through the labels, predicted from the BWF again, to make the labels smoother. The transition probability matrix for MCF is determined based on ground truth data, which specifies how the state (activity label) evolves over time. The reason that MCF labels are more realistic (i.e., they contain less noise and conform better to the actual labels) is that the MCF uses the ratio that BWF has calculated in the previous step (i.e., 0.83) as its threshold, not the threshold defined first (i.e., 0.68), which is higher and makes the MCF less sensitive to local changes. A comprehensive elaboration of these methods is presented in a paper authored by Sabillon et al. [15]. In this paper, the authors have used the smoothed labels of MCF, because the output results of SWF and BWF were not accurate enough (Figure 10).

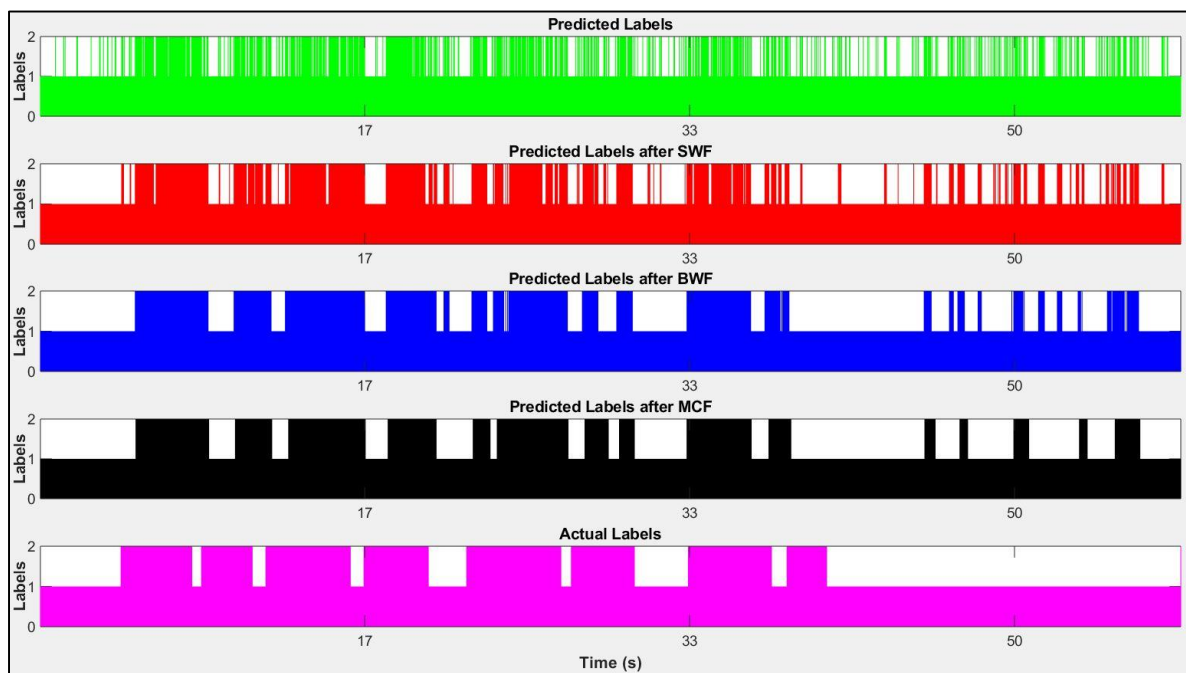


Figure 10. From top to bottom: (1) predicted labels (**green**), (2) predicted labels after Small Window Filtering (SWF) (**red**), (3) predicted labels after Big Window Filtering (BWF) (**blue**), (4) predicted labels after Markov Chain Filtering (MCF) (**black**), (5) actual labels (**magenta**). Black labels represent the activities more realistic than the blue, red, and green ones.

4. Experimental Setup and Results

The aforementioned hybrid system has been evaluated under various conditions, such as different types of equipment and job sites. In this paper, the authors collected and processed data of 10 pieces of construction machines working at three different job sites: (1) jackhammer 305.5E2 (stop, drilling, rotating/moving arm, and moving forward/backward), (2) CAT 259D (stop, arm/shovel movement, moving forward/backward, and turning right/left), (3) SkyJack SJ6826 (maneuvering forward/backward, and raising/lowering), (4) XTREME 842Lift (stop, moving forward/backward, and moving arm), (5) CAT 308E (scraping and moving/rotating arm), (6) CAT 305.5E2 (extending arm and rotating cabin), (7) Dozer 850K (stop, moving forward, and moving backward), (8) Concrete Truck (pouring concrete and moving forward/backward), (9) CAT 938M (stop, moving forward/pushing soil, and moving backward), and (10) CAT 210G Vibrator (stop and vibrating). All of the activities of these machines during certain periods were observed, and the audio and kinematic signals generated were recorded accordingly. The collected audio and kinematic signals were stored as vectors and then exported to a MATLAB platform for further analysis. A Zoom H1 digital handy recorder and an iPhone 6s Plus have been used for data collection purposes. The selected off-the-shelf microphone was located within 50 feet from the desired equipment on a fixed tripod to avoid any movement noises. To record kinematic signals, the iPhone had been placed inside the operator's cabin and mounted to fix its position and prevent any unwanted noise signals from the phone. Figure 11 shows the configuration of the devices in a sample job site. Furthermore, a video camera was utilized for recording all activities of equipment and producing ground truth data. Finally, an equipment horn was used to generate a short distinct sound to properly set the recording beginning point.

The recorded audio and kinematic signals are pre-processed. Then, the aforementioned features are extracted and using PCA, the dimension of the feature set is reduced. In the next step, these features are fused using the feature aggregation method. Next, the aggregated features are fed into the SVM model to determine different activities. Finally, the accuracy of the proposed system has been evaluated by comparing detected labels with actual labels.

To quantify the accuracy of labeling, confusion matrices were completed for each piece of equipment. These matrices illustrate how the predicted labels are precise when compared with their correlated actual labels. Accuracy, defined as the ratio of correctly predicted labels to the total number of labels, has been separately calculated for kinematic signals, audio signals, and fused data in this paper. Table 3 presents the results of calculating confusion matrices for a CAT 259D, with different accuracies for activity detection using kinematic signal, audio signal, and fused data. The value for the kinematic signal is the accuracy level after fusing all six components of acceleration and angular velocity sensors, which are then fused with the audio signal. Similarly, confusion matrices for other types of equipment have been calculated, and their related accuracies are presented in Figure 12.



Figure 11. Configuration of audio and kinematic recording devices in a sample job site.

Table 3. Confusion matrix for CAT 259D.

		Actual Label				Accuracy %
		Stop	Arm/shovel Movement	Moving Forward/Backward	Turning Right/Left	
Predicted Label	Vibration	Stop	41	58	21	87.08
		Arm/shovel movement	19	692	6	
		Moving forward/backward	19	15	227	
		Turning right/left	9	22	1	
Predicted Label	Audio	Stop	11	94	10	74.71
		Arm/shovel movement	3	690	19	
		Moving forward/backward	5	34	164	
		Turning right/left	1	73	44	
Predicted Label	Fused Data	Stop	66	41	10	92.00
		Arm/shovel movement	17	697	1	
		Moving forward/backward	6	4	252	
		Turning right/left	2	11	2	

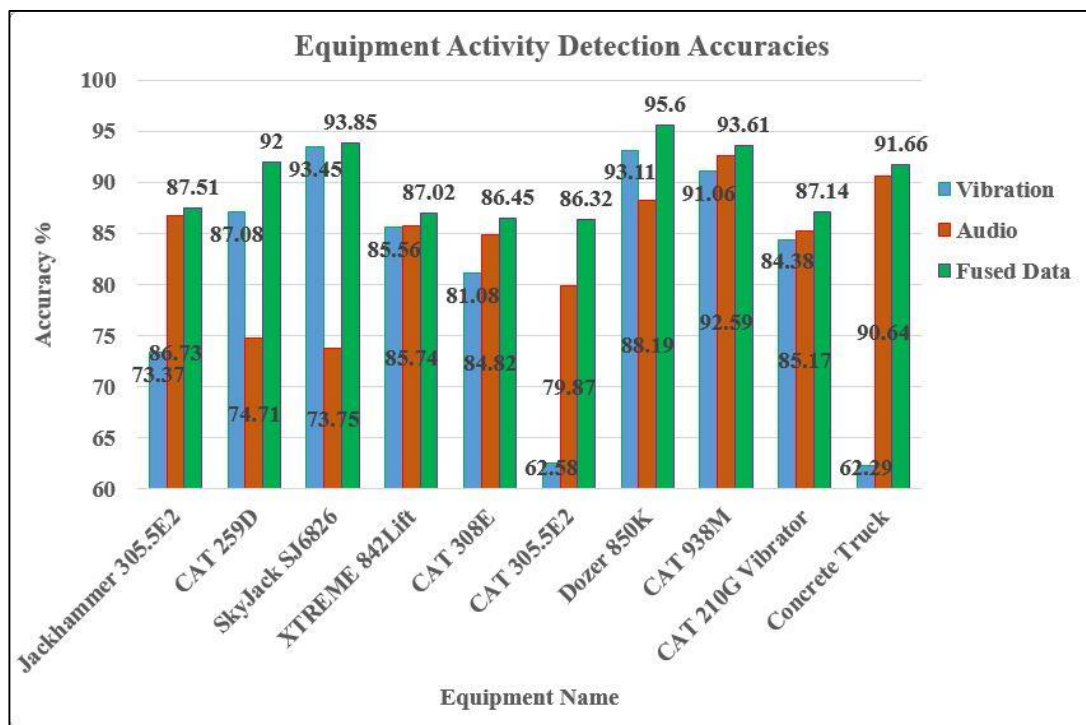


Figure 12. A summary of accuracy rates for activity detection of 10 different pieces of construction equipment.

5. Discussion

5.1. Audio or Kinematic Data?

Careful analysis of the results (presented in Figure 12 and Table 4) verifies the hypothesis that every single type of kinematic or acoustic data would generate more accurate results under certain scenarios. For example, the concrete truck mixer generates very identifiable sound patterns while pouring concrete, so audio is an excellent source of data for this category of equipment's activity classification. At the same time, a kinematic sensor placed inside the cabin is not able to collect reliable data due to the distance and lack of direct connection between the cabin and the discharge chute. Therefore, distance between the kinematic sensor and the signal source is a major factor, which needs to be considered when working with kinematic sensors. Alternatively, for certain types of equipment such as the skyjack SJ6826, kinematic signals are a great source of input data for activity detection purposes because moving forward/backward and raising/lowering the arm does not generate distinct sound, and only their corresponding kinematic signals are detectable. For jackhammer 305.5E2, both data sources are appropriate for activity detection because drilling rocks generates sufficient vibration and sound signals. A similar situation exists for Dozer 850K, where either kinematic or acoustic signals could provide accurate results. The case of the CAT 305.5E2 was challenging, as neither audio nor kinematic data could individually produce accurate results.

To sum up, excavators (e.g., CAT 308E and CAT 305.5E2) generate sufficient sound, which makes audio a better source of input data for activity detection purposes. On the other hand, for bulldozers (e.g., Dozer 850K) kinematic data is a better option, similar for lifts (e.g., skyjack SJ6826 and XTREME 842). For skid-steer loaders (e.g., CAT 259D), both data sources are capable of generating reliable results (Table 4). In Table 4, low, moderate, and high columns show accuracies of 0–75%, 75–85%, and 85–100%, respectively.

Table 4. Classification accuracy.

	Vibration Accuracy			Audio Accuracy			Fused Data Accuracy		
	Low	Moderate	High	Low	Moderate	High	Low	Moderate	High
Jackhammer	✓					✓			✓
305.5E2									
CAT 259D			✓	✓					✓
Skyjack SJ6826			✓	✓					✓
XTREME 842Lift			✓			✓			✓
CAT 308E		✓			✓				✓
CAT 305.5E2	✓				✓				✓
Dozer 850K			✓			✓			✓
CAT 938M			✓			✓			✓
CAT 210G						✓			✓
Vibrator		✓				✓			✓
Concrete Truck	✓					✓			✓

5.2. Role of Data Fusion

As indicated in Table 4 and Figure 12, fusing both audio and vibration data provides a significant improvement in results. Considering the results of implementing the proposed system, it is possible to achieve up to 20% higher accuracy rates by fusing acoustic and kinematic data.

Other than the issues discussed in the previous section, “time synchronization” is an essential cause for obtaining less accurate results. The recorded audio and kinematic signals do not have the same starting points and durations. The synchronization issue has been addressed through a manual procedure, as explained in Section 3.2. The authors intend to develop an automated synchronization procedure as part of their future research plans.

5.3. Applications of Equipment Activity

Recognizing activities of construction equipment might not be useful. In other words, it is the basis for performance monitoring of construction job sites. Construction managers are interested in recognizing equipment activities due to its potential applications for construction performance monitoring. The potential outcomes of activity recognition are as follows:

1. **Maintenance Assessment:** Collecting data from different sensors and recognizing equipment activities can further be used as a platform for monitoring its abnormality or well-being, fuel consumption evaluation, and utilization time and cycle time estimation. Using an automated real-time framework, construction managers are able to continuously monitor activities of equipment using a standalone device or a mobile app. They can be provided with productivity rates and make proper decisions based on the performance of the equipment. For example, they can be notified if an equipment productivity rate is low and make proper decisions such as repairing the equipment, changing the equipment, or even changing the equipment operator.
2. **Environmental Performance Monitoring:** Construction equipment usually releases detrimental smoke, which makes the construction job site unhealthy and unsafe for personnel. Also, it has harmful impacts on the environment. Thus, continuous monitoring of equipment can help construction managers track the emissions of the equipment during its operations and find its potential deficiencies. In other words, engine audio and kinematic data of equipment can be compared with the new models’ data and identify any abnormality to prevent more pollution.

6. Conclusions

In this paper, the authors propose a robust hybrid equipment activity detection system based upon using both audio and kinematic data. The proposed method is validated using ten types of equipment and the obtained results were promising. The authors demonstrated the capabilities of a hybrid system in overcoming limitations of single sources of data and generating more accurate results for a broader

range of construction equipment. Several efficient pre-processing algorithms (i.e., de-noising, removing gravity, removing outliers, and filling missing values) were implemented in this paper to refine the data and improve the overall performance of the proposed system. Considering the computational efficiency of the implemented algorithms, the proposed method is also capable of being used in near real-time scenarios and construction operations can be monitored continuously without the need for human interference. In Table 5, computational time for testing three types of equipment for different types of data is shown. As illustrated here, the computational time is not directly related to the number of activities and is less than one second, which could provide an indication of possible use for near real-time applications.

Furthermore, activity recognition of construction equipment is the first step toward construction operation monitoring and it can help construction managers in different aspects such as productivity analysis, equipment downtime/idle time detection, equipment cycle time estimation, and equipment fuel use control.

Following are some of the key contributions of this research:

1. Construction job sites may vary in types of existing equipment, weather, complexity, etc. Some types of equipment, especially new models, might not generate kinematic signals and this would make it almost impossible to detect activities. Also, distance and inaccessibility of equipment or the presence of sound barriers may hinder the process of recording audios. Rainy, big and crowded job sites might affect the accuracy of each data type and decrease the precision of the detection. In this paper, both audio and kinematic signals have been fused to overcome and cover these limitations.
2. Most of the aforementioned papers utilized a few features to train the machine learning model. In this paper, different types of time-domain and frequency-domain features were selected and evaluated before using in a training model. Also, a dimension reduction method has been implemented on the feature set to decrease the correlation between features and increase the class separability of features' values. Moreover, it decreases computational time of the process which can further be used in real-time.
3. Different types of pre-processing algorithms were implemented in this paper on audio and kinematic signals which refine the data before being used in subsequent steps.

The focus of this research has been on activity analysis of single machines. The more realistic yet challenging scenario occurs when multiple machines operate simultaneously on a job site. For multiple machine recordings, sound and vibration signals from different machines come from different directions. A robust source detection and activity recognition algorithm must respond to sound and kinematic signals from a specific direction and block most of the noise outside the direction of interest. Addressing this important issue would be this research project's future extension. In addition, and as part of future studies, the authors will investigate the use of robust PCA, because it has less sensitivity to outliers and it might increase the accuracy of feature reduction.

Table 5. The computational time of different data types for real-time purposes.

(1) Equipment	(2) Number of Activities	(3) Capturing and Pre-processing Data	Testing Time (s)					
			Predicting Time			Total Time		
			(4) Vibration Data	(5) Audio Data	(6) Fused Data	(3) + (4) Vibration Data	(3) + (5) Audio Data	(3) + (6) Fused Data
Concrete Truck	2	0.918	0.043	0.008	0.046	0.961	0.926	0.964
XTREME 842Lift	3	0.963	0.030	0.020	0.015	0.993	0.983	0.978
CAT 259D	4	0.664	0.191	0.035	0.083	0.855	0.699	0.747

Author Contributions: Conceptualization, B.S. and A.R.; Data curation, B.S. and A.R.; Formal analysis, B.S.; Funding acquisition, A.R.; Investigation, B.S.; Methodology, B.S., A.R., Y.-C.L. and C.R.A.; Project administration, A.R.; Resources, A.R.; Software, B.S.; Supervision, A.R., Y.-C.L. and C.R.A.; Validation, B.S.; Visualization, B.S.; Writing—original draft, B.S.; Writing—review and editing, B.S., A.R., Y.-C.L. and C.R.A.

Funding: This research was funded by the U.S. National Science Foundation (NSF), CMMI-1606034.

Acknowledgments: The authors gratefully acknowledge NSF's support. Any opinions, findings, conclusions, and recommendations expressed in this manuscript are those of the authors and do not reflect the views of the funding agency. The authors also appreciate the assistance of Richard Peterson and Adrian Porras, undergraduate students at the University of Utah, with data collection and audio recordings.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cheng, C.-F.; Rashidi, A.; Davenport, M.A.; Anderson, D.V. Activity analysis of construction equipment using audio signals and support vector machines. *Autom. Constr.* **2017**, *81*, 240–253. [[CrossRef](#)]
2. Cheng, C.-F.; Rashidi, A.; Davenport, M.A.; Anderson, D.V. Evaluation of Software and Hardware Settings for Audio-Based Analysis of Construction Operations. *Int. J. Civ. Eng.* **2019**, *17*, 1469–1480. [[CrossRef](#)]
3. Rezazadeh Azar, E.; McCabe, B. Automated. Visual recognition of dump trucks in construction videos. *J. Comput. Civ. Eng.* **2011**, *26*, 769–781. [[CrossRef](#)]
4. Teizer, J.; Allread, B.S.; Fullerton, C.E.; Hinze, J. Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system. *Autom. Constr.* **2010**, *19*, 630–640. [[CrossRef](#)]
5. Yang, J.; Shi, Z.; Wu, Z. Vision-based action recognition of construction workers using dense trajectories. *Adv. Eng. Inform.* **2016**, *30*, 327–336. [[CrossRef](#)]
6. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2019**, *119*, 3–11. [[CrossRef](#)]
7. Noghabaei, M.; Asadi, K.; Han, K. Virtual Manipulation in an Immersive Virtual Environment: Simulation of Virtual Assembly. In Proceedings of the Computing in Civil Engineering 2019: Visualization, Information Modeling and Simulation, Atlanta, GA, USA, 17–19 June 2019.
8. Asadi, K.; Ramshankar, H.; Noghabaei, M.; Han, K. Real-time Image Localization and Registration with BIM Using Perspective Alignment for Indoor Monitoring of Construction. *J. Comput. Civ. Eng.* **2019**, *33*, 04019031. [[CrossRef](#)]
9. Taghaddos, H.; Mashayekhi, A.; Sherafat, B. Automation of Construction Quantity Take-Off: Using Building Information Modeling (BIM). In Proceedings of the Construction Research Congress 2016, San Juan, PR, USA, 31 May–2 June 2016.
10. Golparvar-Fard, M.; Heydarian, A.; Niebles, J.C. Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers. *Adv. Eng. Inform.* **2013**, *27*, 652–663. [[CrossRef](#)]
11. Gong, J.; Caldas, C.H. Computer vision-based video interpretation model for automated productivity analysis of construction operations. *J. Comput. Civ. Eng.* **2009**, *24*, 252–263. [[CrossRef](#)]
12. Gong, J.; Caldas, C.H.; Gordon, C. Learning and classifying actions of construction workers and equipment using Bag-of-Video-Feature-Words and Bayesian network models. *Adv. Eng. Inform.* **2011**, *25*, 771–782. [[CrossRef](#)]
13. Akhavian, R.; Behzadan, A.H. Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers. *Adv. Eng. Inform.* **2015**, *29*, 867–877. [[CrossRef](#)]
14. Ahn, C.R.; Lee, S.; Pena-Mora, F. Application of low-cost accelerometers for measuring the operational efficiency of a construction equipment fleet. *J. Comput. Civ. Eng.* **2013**, *29*, 04014042. [[CrossRef](#)]
15. Sabillon, C.A.; Rashidi, A.; Samanta, B.; Cheng, C.F.; Davenport, M.A.; Anderson, D.V. A Productivity Forecasting System for Construction Cyclic Operations Using Audio Signals and a Bayesian Approach. In Proceedings of the Construction Research Congress 2018, New Orleans, LA, USA, 2–4 April 2018.
16. Sherafat, B.; Rashidi, A.; Lee, Y.C.; Ahn, C.R. Automated Activity Recognition of Construction Equipment Using a Data Fusion Approach. In Proceedings of the Computing in Civil Engineering 2019: Data, Sensing, and Analytics, Atlanta, GA, USA, 17–19 June 2019.

17. Pires, I.M.; Santos, R.; Pombo, N.; Garcia, N.M.; Florez-Revuelta, F.; Spinsante, S.; Goleva, R.; Zdravevski, E. Recognition of Activities of Daily Living Based on Environmental Analyses Using Audio Fingerprinting Techniques: A Systematic Review. *Sensors* **2018**, *18*, 160. [[CrossRef](#)] [[PubMed](#)]
18. Guo, G.; Li, S.Z. Content-based audio classification and retrieval by support vector machines. *IEEE Trans. Neural Netw.* **2003**, *14*, 209–215. [[PubMed](#)]
19. Rossi, M.; Feese, S.; Amft, O.; Braune, N.; Martis, S.; Tröster, G. AmbientSense: A real-time ambient sound recognition system for smartphones. In Proceedings of the 2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), San Diego, CA, USA, 18–22 March 2013.
20. Jayalakshmi, S.L.; Chandrakala, S.; Nedunchelian, R. Global statistical features-based approach for Acoustic Event Detection. *Appl. Acoust.* **2018**, *139*, 113–118. [[CrossRef](#)]
21. Saki, F.; Sehgal, A.; Panahi, I.; Kehtarnavaz, N. Smartphone-based real-time classification of noise signals using subband features and random forest classifier. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016.
22. Saki, F.; Kehtarnavaz, N. Real-time hierarchical classification of sound signals for hearing improvement devices. *Appl. Acoust.* **2018**, *132*, 26–32. [[CrossRef](#)]
23. Lu, L.; Zhang, H.J.; Jiang, H. Content analysis for audio classification and segmentation. *IEEE Trans. Speech Audio Process.* **2002**, *10*, 504–516. [[CrossRef](#)]
24. Chathuranga, Y.M.D.; Jayaratne, K.L. Automatic music genre classification of audio signals with machine learning approaches. *GSTF J. Comput. (JoC)* **2018**, *3*, 13.
25. Gerhard, D. *Audio Signal Classification: History and Current Techniques*; Technical Report TR-CS 2003–07; Department of Computer Science, University of Regina: Regina, SK, Canada, 2003; ISBN 0 7731 0456 9.
26. Virtanen, T.; Plumbley, M.D.; Ellis, D. Introduction to sound scene and event analysis. In *Computational Analysis of Sound Scenes and Events*; Springer International Publishing: New York City, NY, USA, 2018; pp. 3–12.
27. Leonardo, R.; Barandas, M.; Gamboa, H. A framework for infrastructure-free indoor localization based on pervasive sound analysis. *IEEE Sens. J.* **2018**, *18*, 4136–4144. [[CrossRef](#)]
28. Hinton, G.; Mohamed, A.-R.; Jaitly, N.; Vanhoucke, V.; Kingsbury, B.; Deng, L.; Yu, D.; Dahl, G.; Senior, A.V.; Nguyen, P.; et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal. Process. Mag.* **2012**, *29*, 82–97. [[CrossRef](#)]
29. McLoughlin, I.V.; Zhang, H.; Xie, Z.; Song, Y.; Xiao, W. Robust sound event classification using deep neural networks. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 540–552. [[CrossRef](#)]
30. Graves, A.; Mohamed, A.R.; Hinton, G. Speech recognition with deep recurrent neural networks. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013.
31. Hwang, K.; Lee, S.Y. Environmental audio scene and activity recognition through mobile-based crowdsourcing. *IEEE Trans. Consum. Electronics* **2012**, *58*, 700–705. [[CrossRef](#)]
32. Zhang, T.; Lee, Y.C.; Scarpiniti, M.; Uncini, A. A Supervised Machine Learning-Based Sound Identification for Construction Activity Monitoring and Performance Evaluation. In Proceedings of the Construction Research Congress, New Orleans, LA, USA, 2–4 April 2018.
33. Sherafat, B.; Rashidi, A.; Song, S. A Software-Based Approach for Acoustical Modeling of Construction Job Sites with Multiple Operational Machines. In Proceedings of the Construction Research Congress, Tempe, AZ, USA, 8–10 March 2020. (Accepted).
34. Cheng, C.F.; Anderson, D.V.; Davenport, M.A.; Rashidi, A. Audio Classification Based on Weakly Labeled Data. In Proceedings of the IEEE Statistical Signal Processing Workshop (SSP), Freiburg, Germany, 10–13 June 2018.
35. Bao, L.; Intille, S.S. April. Activity recognition from user-annotated acceleration data. In Proceedings of the International Conference on Pervasive Computing, Linz/Vienna, Austria, 21–23 April 2004.
36. Wang, N.; Ambikairajah, E.; Lovell, N.H.; Celler, B.G. Accelerometry based classification of walking patterns using time-frequency analysis. In Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, 22–26 August 2007.
37. Cho, Y.; Nam, Y.; Choi, Y.J.; Cho, W.D. Smart Buckle: Human activity recognition using a 3-axis accelerometer and a wearable camera. In Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments, Breckenridge, CO, USA, 17 June 2008.

38. Altun, K.; Barshan, B.; Tunçel, O. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognit.* **2010**, *43*, 3605–3620. [[CrossRef](#)]
39. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity recognition using cell phone accelerometers. *ACM SigKDD Explor. Newsl.* **2011**, *12*, 74. [[CrossRef](#)]
40. Weng, S.; Xiang, L.; Tang, W.; Yang, H.; Zheng, L.; Lu, H.; Zheng, H. A low power and high accuracy MEMS sensor based activity recognition algorithm. In Proceedings of the 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Belfast, UK, 2–5 November 2014.
41. Gao, L.; Bourke, A.K.; Nelson, J. Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Med. Eng. Phys.* **2014**, *36*, 779–785. [[CrossRef](#)] [[PubMed](#)]
42. Bersch, S.D.; Azzi, D.; Khusainov, R.; Achumba, I.E.; Ries, J. Sensor data acquisition and processing parameters for human activity classification. *Sensors* **2014**, *14*, 4239–4270. [[CrossRef](#)]
43. Ignatov, A.D.; Strijov, V.V. Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer. *Multimed. Tools Appl.* **2016**, *75*, 7257–7270. [[CrossRef](#)]
44. Anwary, A.R.; Yu, H.; Vassallo, M. Optimal foot location for placing wearable IMU sensors and automatic feature extraction for gait analysis. *IEEE Sens. J.* **2017**, *18*, 2555–2567. [[CrossRef](#)]
45. Lee, K.; Kwan, M.P. Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results. *Comput. Environ. Urban. Syst.* **2018**, *67*, 124–131. [[CrossRef](#)]
46. Ahn, C.R.; Lee, S.; Peña-Mora, F. Monitoring system for operational efficiency and environmental performance of construction operations using vibration signal analysis. In Proceedings of the Construction Research Congress, West Lafayette, IN, USA, 21–23 May 2012.
47. Akhavian, R.; Behzadan, A.H. Remote monitoring of dynamic construction processes using automated equipment tracking. In Proceedings of the Construction Research Congress, West Lafayette, IN, USA, 21–23 May 2012.
48. Akhavian, R.; Behzadan, A.H. Knowledge-based simulation modeling of construction fleet operations using multimodal-process data mining. *J. Constr. Eng. Manag.* **2013**, *139*, 4013021. [[CrossRef](#)]
49. Akhavian, R.; Behzadan, A.H. Client-server interaction knowledge discovery for operations-level construction simulation using process data. In Proceedings of the Construction Research Congress 2014, Atlanta, GA, USA, 19–21 May 2014.
50. Kim, H.; Ahn, C.R.; Engelhaupt, D.; Lee, S. Application of dynamic time warping to the recognition of mixed equipment activities in cycle time measurement. *Autom. Constr.* **2018**, *87*, 225–234. [[CrossRef](#)]
51. Garcia-Ceja, E.; Galván-Tejada, C.E.; Brena, R. Multi-view stacking for activity recognition with sound and accelerometer data. *Inf. Fusion* **2018**, *40*, 45–56. [[CrossRef](#)]
52. Rangachari, S.; Loizou, P.C. A noise-estimation algorithm for highly non-stationary environments. *Speech Commun.* **2006**, *48*, 220–231. [[CrossRef](#)]
53. Lau, H.Y.; Tong, K.Y.; Zhu, H. Support vector machine for classification of walking conditions using miniature kinematic sensors. *Med. Boil. Eng. Comput.* **2008**, *46*, 563–573. [[CrossRef](#)] [[PubMed](#)]
54. Fujiki, Y. iPhone as a physical activity measurement platform. In Proceedings of the CHI'10 Extended Abstracts on Human Factors in Computing Systems 2010, Atlanta, GA, USA, 10–15 April 2010; pp. 4315–4320.
55. Bayat, A.; Pomplun, M.; Tran, D.A. A study on human activity recognition using accelerometer data from smartphones. *Procedia Comput. Sci.* **2014**, *34*, 450–457. [[CrossRef](#)]
56. Gerami, A. *Introduction to Digital Communications*, 1st ed.; Academic Press: Cambridge, MA, USA, 2015; pp. 217–264.
57. Zhang, M.; Cao, T.; Zhao, X. Applying Sensor-Based Technology to Improve Construction Safety Management. *Sensors* **2017**, *17*, 1841. [[CrossRef](#)] [[PubMed](#)]
58. Park, M.W.; Koch, C.; Brilakis, I. Three-dimensional tracking of construction resources using an on-site camera system. *J. Comput. Civ. Eng.* **2011**, *26*, 541–549. [[CrossRef](#)]
59. Kozhisseri, S.; Bikdash, M. Spectral features for the classification of civilian vehicles using acoustic sensors. In Proceedings of the 2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems, Nashville, TN, USA, 30 March–2 April 2009.
60. Wiczorkowska, A.; Kubera, E.; Słowik, T.; Skrzypiec, K. Spectral features for audio-based vehicle and engine classification. *J. Intell. Inf. Syst.* **2018**, *50*, 265–290. [[CrossRef](#)]

61. Wei, J.; Vongsy, K.; Mendoza-Schrock, O.; Liu, C.H. Vehicle Engine Classification Using Spectral Tone-Pitch Vibration Indexing and Neural Network. *Int. J. Monit. Surveill. Technol. Res. (IJMSTR)* **2014**, *2*, 31–49. [[CrossRef](#)]
62. Padmavathi, G.; Shanmugapriya, D.; Kalaivani, M. Neural network approaches and MSPCA in vehicle acoustic signal classification using wireless sensor networks. In Proceedings of the 2010 International Conference on Communication Control and Computing Technologies, Ramanathapuram, India, 7–9 October 2010.
63. Walse, K.H.; Dharaskar, R.V.; Thakare, V.M. Pca based optimal ann classifiers for human activity recognition using mobile sensors data. In *Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems*; Springer International Publishing: New York City, NY, USA, 2016; pp. 429–436.
64. Mishra, D.; Dash, R.; Rath, A.K.; Acharya, M. Feature Selection in Gene Expression Data Using Principal Component Analysis and Rough Set Theory. In *Software Tools and Algorithms for Biological Systems. Advances in Experimental Medicine and Biology*; Arabnia, H., Tran, Q.N., Eds.; Springer: New York City, NY, USA, 2011; Volume 696.
65. Jain, A.; Nandakumar, K.; Ross, A. Score normalization in multimodal biometric systems. *Pattern Recognit.* **2005**, *38*, 2270–2285. [[CrossRef](#)]
66. Guilloux, J.P.; Seney, M.; Edgar, N.; Sibille, E. Integrated behavioral z-scoring increases the sensitivity and reliability of behavioral phenotyping in mice: Relevance to emotionality and sex. *J. Neurosci. Methods* **2011**, *197*, 21–31. [[CrossRef](#)] [[PubMed](#)]
67. Kaiser, H.F. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* **1960**, *20*, 141–151. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).