Project 1852 | December 2019

SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Automated Measurement of Heavy Equipment Greenhouse Gas Emission: The Case of Road/Bridge Construction and Maintenance

Reza Akhavian, PhD

CSUTC California State University Transportation Consortium

CAL STATE EAST BAY

MINETA TRANSPORTATION INSTITUTE

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the four-university. MTI leads the four-university California State University Transportation Consortium funded by the State of California through Senate Bill 1.

MTI's transportation policy work is centered on three primary responsibilities:

Research

MTI works to provide policy-oriented research for all levels of government and the private sector to foster the development of optimum surface transportation systems. Research areas include: bicycle and pedestrian issues; financing public and private sector transportation improvements; intermodal connectivity and integration; safety and security of transportation systems; sustainability of transportation systems; transportation / land use / environment; and transportation planning and policy development. Certified Research Associates conduct the research. Certification requires an advanced degree, generally a Ph.D., a record of academic publications, and professional references. Research projects culminate in a peer-reviewed publication, available on TransWeb, the MTI website (http://transweb.sjsu.edu).

Education

The Institute supports education programs for students seeking a career in the development and operation of surface transportation systems. MTI, through San José State University, offers an AACSB-accredited Master of Science in Transportation Management and graduate certificates in Transportation Management, Transportation Security, and High-Speed Rail Management that serve to prepare the nation's transportation managers for the 21st century. With the active assistance of the California Department of Transportation (Caltrans), MTI delivers its classes over a state-of-the-art videoconference network throughout the state of California and via webcasting beyond, allowing working transportation professionals to pursue an advanced degree regardless of their location. To meet the needs of employers seeking a diverse workforce, MTI's education program promotes enrollment to under-represented groups.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and journals and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation- related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

REPORT 19-34

AUTOMATED MEASUREMENT OF HEAVY EQUIPMENT GREENHOUSE GAS EMISSION: THE CASE OF ROAD/ BRIDGE CONSTRUCTION AND MAINTENANCE

Reza Akhavian, PhD

December 2019

A publication of

Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 19-34	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Automated Measurement of Heavy Equipment Greenhouse Gas Emission: The case of Road/Bridge Construction and Maintenance		**5. Report Date** December 2019 **6. Performing Organization Code**
7. Authors Reza Akhavian, PhD https://orcid.org/ 0000-0001-9691-8016		**8. Performing Organization Report** CA-MTI-1852
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219		**10. Work Unit No.** **11. Contract or Grant No.** ZSB12017-SJAUX
12. Sponsoring Agency Name and Address State of California SB1 2017/2018 Trustees of the California State University Sponsored Programs Administration 401 Golden Shore, 5th Floor Long Beach, CA 90802		**13. Type of Report and Period Covered** Final Report **14. Sponsoring Agency Code**
15. Supplemental Notes 10.31979/mti.2019.1852		

16. Abstract

Road/bridge construction and maintenance projects are major contributors to greenhouse gas (GHG) emissions such as carbon dioxide (CO2), mainly due to extensive use of heavy-duty diesel construction equipment and large-scale earthworks and earthmoving operations. Heavy equipment is a costly resource and its underutilization could result in significant budget overruns. A practical way to cut emissions is to reduce the time equipment spends doing non-value-added activities and/or idling. Recent research into the monitoring of automated equipment using sensors and Internet-of-Things (IoT) frameworks have leveraged machine learning algorithms to predict the behavior of tracked entities.

In this project, end-to-end deep learning models were developed that can learn to accurately classify the activities of construction equipment based on vibration patterns picked up by accelerometers attached to the equipment.

Data was collected from two types of real-world construction equipment, both used extensively in road/bridge construction and maintenance projects: excavators and vibratory rollers. The validation accuracies of the developed models were tested of three different deep learning models: a baseline convolutional neural network (CNN); a hybrid convolutional and recurrent long short-term memory neural network (LSTM); and a temporal convolutional network (TCN). Results indicated that the TCN model had the best performance, the LSTM model had the second-best performance, and the CNN model had the worst performance. The TCN model had over 83% validation accuracy in recognizing activities.

Using deep learning methodologies can significantly increase emission estimation accuracy for heavy equipment and help decision-makers to reliably evaluate the environmental impact of heavy civil and infrastructure projects. Reducing the carbon footprint and fuel use of heavy equipment in road/bridge projects have direct and indirect impacts on health and the economy. Public infrastructure projects can leverage the proposed system to reduce the environmental cost of infrastructure project.

17. Key Words road construction, ANNs (artificial neural networks), emissions, construction equipment, greenhouse gases	**18. Distribution Statement** No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 27	**22. Price**

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2019.1852

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu

ACKNOWLEDGMENTS

The author would like to thank Guerra Construction Group and Kiewit Construction for allowing access to their job sites for data collection, as well as student assistants Mr. Carlos Hernandez and Mr. Trevor Slaton.

The authors thank Editing Press, for editorial services, as well as MTI staff, including Executive Director Karen Philbrick, PhD; Deputy Executive Director Hilary Nixon, PhD; Research Support Assistant Anket Sah; Graphic Designer Alverina Eka Weinardy and Executive Administrative Assistant Jill Carter.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Monitoring construction resources used in road/bridge projects, such as heavy equipment, enables not only improvements in productivity but also increased knowledge of emissions produced as a result of fuel consumption. Previous studies conducted by the United States Environmental Protection Agency (EPA) have demonstrated that heavy-duty construction equipment is one of the major contributors of emissions from diesel engines. Diesel engine emissions contain large amounts of carbon monoxide (CO), nitrogen oxides (NOx), hydrocarbons (HC), and particulate matter (PM), all of which have direct negative impacts on human health. A practical way to cut emissions is to reduce the time that construction equipment spends doing non-value-adding activities and/or idling. Recent research in automated equipment monitoring using sensors and Internet-of-Things (IoT) frameworks have leveraged machine learning algorithms to predict the behavior of tracked entities. Previous methodologies, however, depended on manual feature engineering, and were therefore not completely conducive to fully automated, generalizable applications. The advent of deep learning models not only automated the feature extraction step, but also resulted in higher accuracies compared to the performance of traditional and shallow machine-learning methods.

In this project, end-to-end deep learning models were developed that can learn to accurately classify the activities of construction equipment based on vibration patterns picked up by accelerometers attached to the equipment. Additionally, relationships were studied between the equipment activities and the emissions that they generate.

Data was collected from two types of real-world construction equipment, both used extensively in road/bridge construction and maintenance projects: excavators and vibratory rollers. The validation accuracies of the developed models were tested of three different deep learning models: a baseline convolutional neural network (CNN); a hybrid convolutional and recurrent long short-term memory neural network (LSTM); and a temporal convolutional network (TCN). Results indicated that the TCN model had the best performance, the LSTM model had the second-best performance, and the CNN model had the worst performance. The TCN model had over 83% validation accuracy in recognizing activities.

I. INTRODUCTION

Road/bridge construction and maintenance projects are major contributors to greenhouse gas (GHG) emissions such as carbon dioxide (CO2). This is mainly because of the extensive use of heavy-duty diesel (HDD) construction equipment,[1] as well as large-scale earthworks and earthmoving operations involved in such projects.[2] A number of recent studies have highlighted the need for estimation of construction equipment idling time, for purposes such as cost estimation, fuel use and emission estimation, data-driven modeling and simulation.[3,4] However, there has not been much research on the prospect, for sustainability analysis, of automated idling detection and idle time estimation. Diesel engine emissions contain large amounts of carbon monoxide (CO), nitrogen oxides (NOx), hydrocarbons (HC), and particulate matter (PM), all of which have a direct negative impact on human health[5]. A practical way to cut emissions is to reduce the time that construction equipment spends doing non-value-adding activities and/or idling. Research indicates that although using newer equipment, using well-maintained equipment, and using clean fuels can improve exhaust emissions, reducing engine idling time and enhancing equipment operating efficiencies results in even better outcomes.[6],[7] When infrastructure projects are located within densely-populated areas, they can require additional logistics tasks that can raise equipment idling rates as high as 70%.[8]

Traditionally, construction equipment emissions have been measured manually using steady-state engine dynamometer tests.[9] However, manual measurements are error-prone and labor-intensive. Therefore, automated identification of the activities of construction resources has been the subject of many recent studies. A common goal of these studies has been the development of an Internet-of-Things (IoT) framework that uses machine learning techniques to distinguish different activities performed by construction workers and/or construction equipment, based on data collected from various sensors. Developing and validating an accurate activity-recognition framework is a first step toward building a system that reliably monitors productivity and predicts greenhouse gas (GHG) emissions. Heavy equipment generates distinct vibration patterns while performing different kinds of task, that can be picked up by accelerometers attached to the equipment. Readings from these accelerometers can then be analyzed, providing an un-intrusive and highly accurate activity-recognition system.[10]

In this project, end-to-end deep learning models were developed that can learn to accurately classify the activities of construction equipment based on vibration patterns picked up by accelerometers attached to the equipment. Additionally, relationships between the equipment activities and the emissions that they generate were studied.

This study differs from previous work in the area of construction equipment activity recognition using accelerometer sensors because of its use of deep neural network architectures. Using deep learning techniques, higher accuracies are achieved compared to classical machine learning results with less manual effort spent in system design and feature selection. The framework proposed in this research consists of a series of steps: data collection; data processing; data segmentation using sliding windows; and classification of the activities at each time step in the data, using deep learning that are described in detail in this report.

II. DATA COLLECTION

In order to ensure its practicality, the developed framework was applied to actual construction equipment performing real work. Data collection sessions were not conducted in a controlled environment, meaning that the equipment operators were asked to continue their projects' tasks as scheduled. Data were collected in two different session. In session 1, data was collected from a *BOMAG BW 145PDH-3 single drum vibratory roller* as well as a *CAT 328D crawler excavator*. In session 2, data was collected from a *CAT 305D CR excavator*. Session 1 focused on inertial (i.e., accelerometer) data collection from the construction equipment in order to build the underlying activity recognition framework. In session 2, in addition to the accelerometer readings, emissions data for various gases were collected using a portable emission measurement system (PEMS) in order to investigate the activity-emission relationship. To distinguish between the two excavators throughout this paper, the one that was subject to data collection in the first session is referred to as Excavator 1 and the one used in the second session is referred to as Excavator 2.

Activities of the equipment in both sessions were video-recorded for later data annotation and model verification. In each data collection session, two Noraxon-manufactured MyoMotion 684 accelerometer sensors were attached to each vehicle, on articulated parts. A signal receiver antenna connected to a laptop on the jobsite was used to log data in real-time. The software module included with the sensor kit was used for data pre-processing tasks, such as automated synchronization between the sensor data and video recordings, and manual labeling of the activities. Figure 1 shows the data collection station with the laptop and the receiver, as well as the sensor placements for the vibratory roller data collection. One of the two sensors was attached on the cabin dashboard close to the steering wheel; the other was attached on the roller's support arm.

Figure 1. (A) Data Collection Station with the (1) Sensors, (2) Receiver Antenna, and (3) Webcam for Synchronous Video, and (B) Sensor Installation on the Vibratory Roller Body

For the data collected from Excavator 1, one sensor was placed on the cabin dashboard and the other was placed on the excavator arm, very closed to the bucket. Figure 2 shows the sensors' placements on the excavator body.

Figure 2. Sensor Installation on Excavator 1 Body

While the roller and Excavator 1 performed their activities, readings were sampled at a rate of 100 samples per second from two 3-axis accelerometers mounted at two different locations on each machine. These activities generated six channels of 116,536 sensor readings over a period of 20 minutes for the roller and six channels of 173,600 sensor readings over a period of nearly 30 minutes for the excavator.

For Excavator 2, in total, 377,808 accelerometer readings were collected at a sampling rate of 100 samples per second. Because the PEMS, for collecting emissions data, operated at a sampling rate of 1 sample per second, its readings were upsampled to 100 samples per second, to match those of the accelerometers. Figure 3 shows the sensors' placements on the excavator body.

Figure 3. Sensor Installation on Excavator 2 Body

III. METHODOLOGY

This project was undertaken in two phases. In Phase 1, the activity-recognition framework was developed using the data collected from the vibratory roller and Excavator 1. In Phase 2, the developed framework was further revised and improved, and emission data was also incorporated, using the data collected from Excavator 2.

PHASE 1

Previous research used a novel deep, convolutional, LSTM, recurrent neural network architecture called *DeepConvLSTM*, for the task of human gesture recognition after training on multimodal sensor data, and found that this architecture outperformed competing non-recurrent networks.[11] The presented study aimed to test whether the success of combining convolutional layers with long short-term memory (LSTM) layers translates to equipment activity recognition as well. Figure 4 shows an overview of the approach developed here.

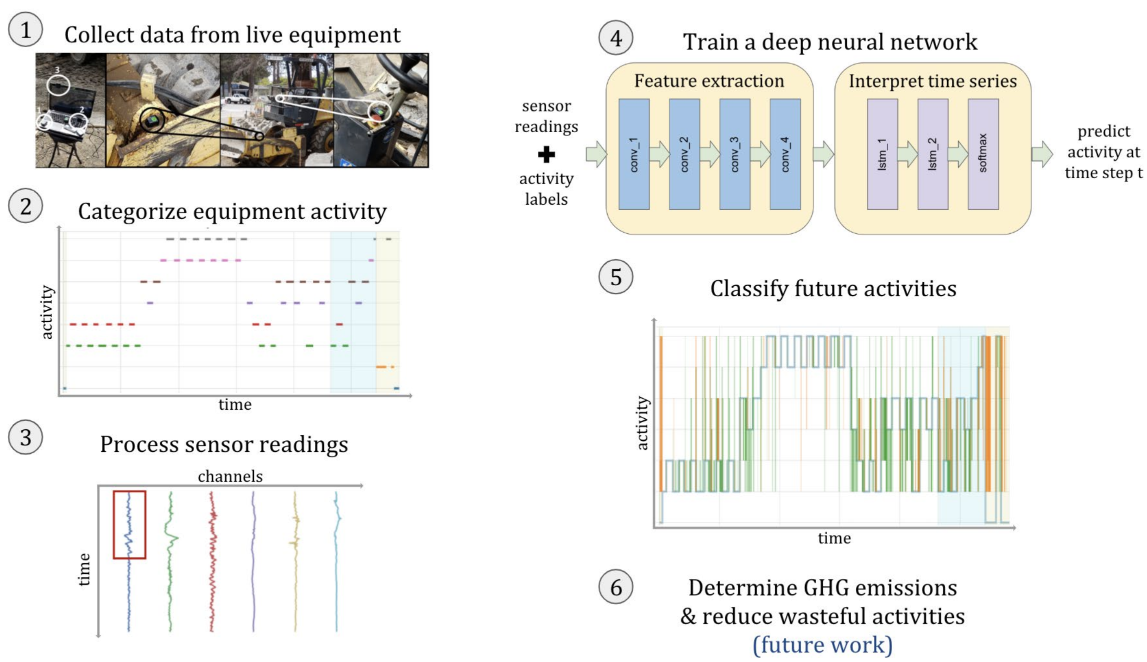

Figure 4. An Overview of the Developed Methodology

Data Analysis

Models were trained and validated on disjoint subsets of the data collected from the roller and the Excavator 1. Validating models on data not seen during training provides a test of the models' real-world predictive power. Across the collected data, the final 20–28% of sensor readings were set aside for validation; the rest of the data were used for training. For the roller data, the first 92,728 samples were used for training, and the remaining samples formed the validation set. This split was chosen so as to maximize the similarity between the activity label distributions of the validation set and the activity label distributions of the training set, while maintaining the correlations between time-adjacent samples that are critical to the problem. Additionally, a small number of samples were dropped from the extreme ends of the data sets to exclude the activities Idle and Off, which were rare in the collected data, from further consideration. As a result, the first 1,040 samples were dropped, as were the last 8,017 samples. In addition to the full problem with six activity classes presented below in Figure 5, two subproblems were also studied, by merging activity classes and re-choosing training and validation sets: the subproblem of distinguishing forward motion from backward motion, and the subproblem of distinguishing activities related to the three vibration modes. The validation set is highlighted in blue, and regions of the data not considered are highlighted in yellow.

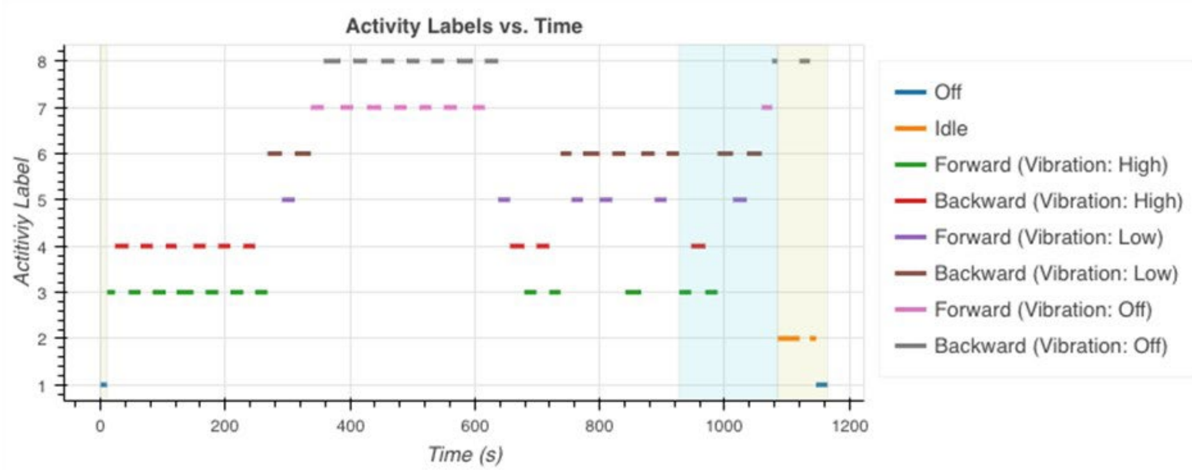

Figure 5. Activity data vs. time for the roller data (validation set in blue; dropped set in yellow)

For Excavator 1, the first 125,165 samples were used for training, and the remaining samples were used for validation. Figure 6 shows the data used for the Excavator 1 tests. Transitions between activities in this dataset were much more frequent than in the roller dataset. No samples were dropped; however, the samples whose activities were labeled “Various” were treated with caution. Two models were trained for Excavator 1: the first model’s training included all data points, while the second model’s training excluded the frames labeled “Various”. The first model was evaluated using all the validation data points, while the second was evaluated using the subset of validation frames excluding those labeled “Various”. This second model was also evaluated on a subset of the validation data points excluding both those labeled “Various” and the first 14,335 labeled “Idle,” in order to better balance the modified class distribution. The model was able to identify the Idle activity with nearly perfect accuracy, so the rebalanced scenario posed a more realistic challenge.

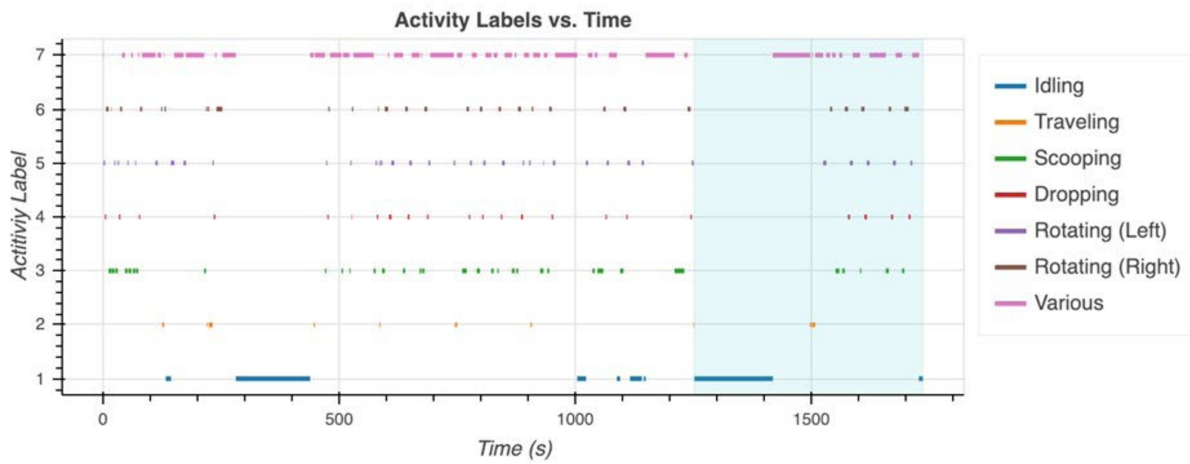

Figure 6. Activity data vs. time for the excavator data (validation set in blue)

In order to learn patterns with strong predictive value, it is critical that models are able to observe the sensor reading at each time step within a larger context of recent sensor readings. To facilitate this, a sliding window length of 200 samples with a step size of 1 sample was used, segmenting the data into overlapping frames corresponding to 2 seconds of activity each. The activity label at the last sample in each frame was used as the label of the frame. This setup formulates the activity-recognition problem as the task of predicting the activity label at each time step in the data series, given the 199 immediately-preceding accelerometer readings. Smaller window sizes could be useful in real-time monitoring applications where a lead-time of 2 seconds is considered too slow; larger window sizes could be useful by providing greater context to each data point when there is greater allowance for time and computational complexity. For this problem, it was only necessary that the window size be large enough to provide adequate context for each sample.

Each accelerometer sensor provides output in three channels, each representing the acceleration component along one dimension (x, y, z), for a total of six sensor channels per piece of equipment. During the sliding-window segmentation process, each sensor channel's output was normalized to the range [0, 1], and stacked horizontally so that each frame contained the time steps on its vertical axis the readings from the six sensor channels on its horizontal axis. Figure 7 depicts an example data frame below, representing one window. The red box—not drawn to scale—represents a filter (0.03 seconds tall, 1 channel wide) that slides across the frame while computing convolutional features

Each signal is labeled by its axis of acceleration (x, y, z) and subscripted with the number of the sensor to which it belongs. The sliding-window segmentation process was applied to the training and validation subsets of the data independently, to prevent validation from leaking into the training data when the frames overlap at the boundary.

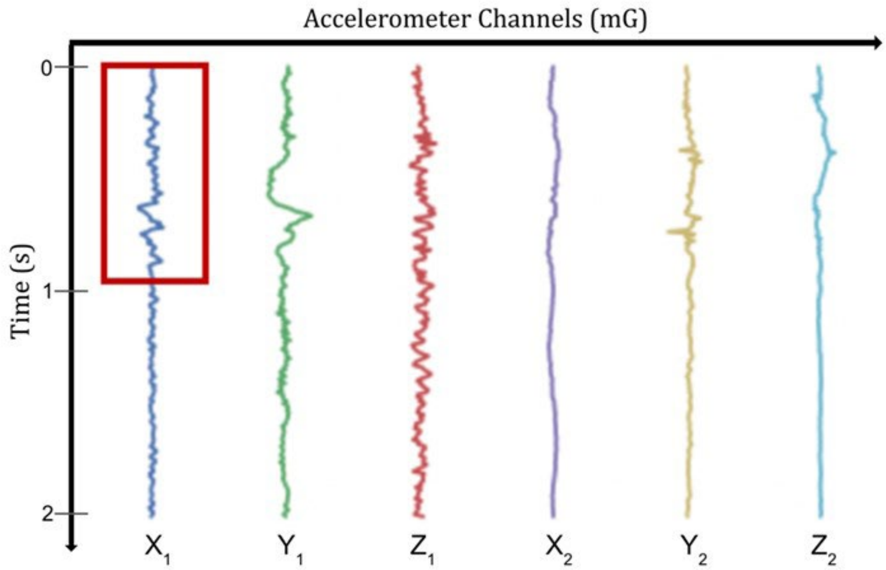

Figure 7. A 2-second frame computed by sliding window

PHASE 2

Data Analysis

For Excavator 2, 377,808 accelerometer readings were collected at a sampling rate of 100 samples per second. Because the PEMS operated at a sampling rate of 1 sample per second, its readings were upsampled to 100 samples per second to match those of the accelerometers. The first 324,579 readings (85.9%) were used as training data while the remaining 53,229 readings (14.1%) were used for validation of the results. This split was chosen so as to ensure similar activity distributions in the training and validation sets. The data set for Excavator 2 is shown in Figure 8 below.

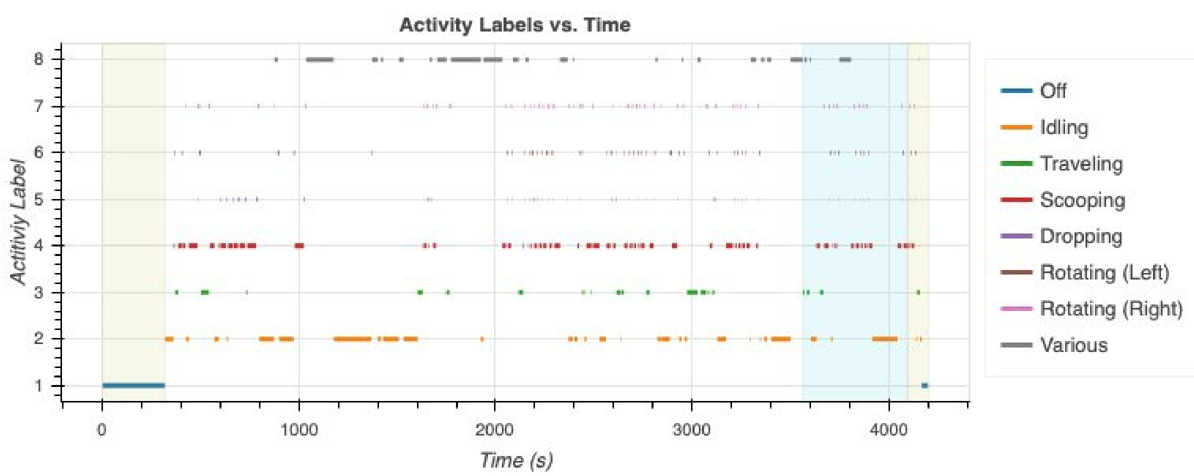

Figure 8. Labeled activities of the CAT 305D CR excavator (validation set in blue, discarded set in yellow)

The emissions signals collected and studied include carbon monoxide (CO), nitrogen oxides (NOx), and carbon dioxide (CO_2). The CO_2 measurement was performed using infrared (IR). Because the carbon dioxide emissions were much larger, they are reported on a percentage scale, while the other signals are reported in parts-per-million (ppm). Figure 9 plots the emission signals vs. time below.

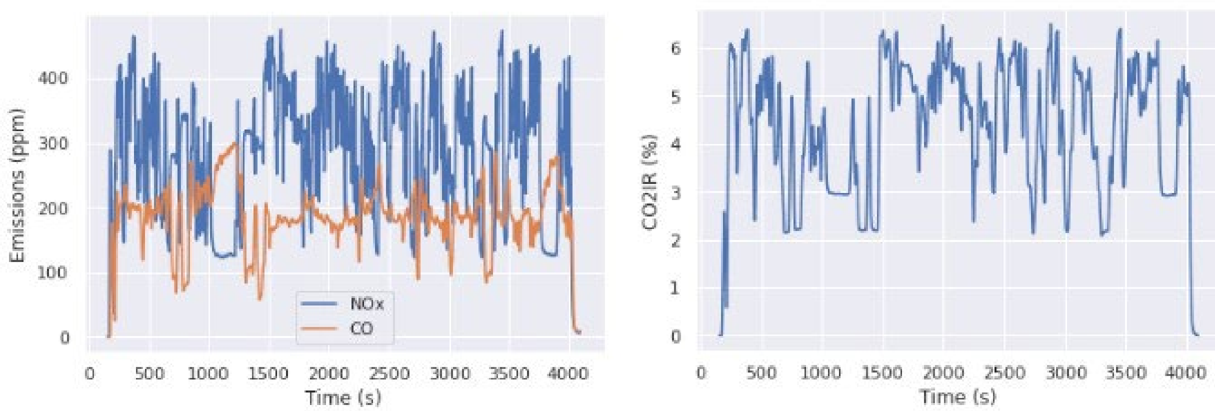

Figure 9. CO, NOx, and CO2 emissions vs. time

The same data processing techniques were applied as in Phase 1 for activity recognition. That is, the readings in each accelerometer sensor channel were normalized to fall into the range [0, 1] and segmented into data frames of 200 time steps by 6 sensor channels each (two sensors in x, y, and z directions each), using an overlapping sliding-window process. Each frame was labeled according to the activity at the last time step, formulating the activity-recognition problem as the task of predicting the activity label at the 200th time step based on the immediately preceding 199 accelerometer readings.

For each emissions signal considered, the readings were separated by activity and plotted as density histograms, estimating their true distributions. The training and validation subsets of the data were plotted separately. Please refer to Linking Activities to Emission in Section IV. Results for detailed description of the histograms.

In Phase 1, the authors studied *BaselineCNN* (“Baseline Convolutional Neural Net”) and *DeepConvLSTM* (“Deep Convolutional Long Short-Term Memory Neural Net”), two models adapted for construction equipment activity recognition based on models of the same names originally developed for human activity recognition by Ordóñez and Roggen (2016). Both *BaselineCNN* and *DeepConvLSTM* begin with four layers of convolutional filters meant to automatically extract features from the accelerometer time series. *BaselineCNN* then uses two fully-connected layers to use these extracted features and make a classification, while *DeepConvLSTM* uses two long short-term memory (LSTM) layers to use the extracted features and make a classification. LSTMs are a particularly popular and high-performing kind of recurrent neural network (RNN), a broad class of neural networks distinguished by the presence of loops; unlike feedforward neural networks, a RNNs state is able to act as a sort of memory influencing future states, allowing the RNN to recognize not only individual inputs but sequences of inputs.

In Phase 2, Temporal convolutional networks (TCNs) was also investigated in addition to the *BaselineCNN* and *DeepConvLSTM*. TCNs are another kind of network designed to deal with sequence data. Traditional convolutional networks (CNNs) are suited to extracting locally correlated features, but not suited to interpreting features that are distant from each other in space or time. This is because the receptive field of a convolutional network scales linearly with its number of layers. In order to achieve a larger receptive field, one that scales exponentially with the number of convolutional layers, van den Oord et al. (2019)[12] applied the concept of dilated convolutions to CNNs. Equation 1 describes how to convolve a 1D filter K, of width W, and dilation factor d, with discrete input signal X(τ):

$$(K *_d X)(\tau) = \sum_{i=1}^{W} K(i)X(\tau - d \cdot i)K *_d X)(\tau) = \sum_{i=1}^{W} K(i)X(\tau - d \cdot i) \qquad (1)$$

When d=1, this is just the usual convolution operation. The factor d scales the amount of space between the adjacent samples of the input signal that get multiplied by the corresponding entries in the filter.

IV. RESULTS

PHASE 1

The models' parameters were optimized over five epochs using batched gradient descent with a batch size of 100 frames and the Adam optimizer with a learning rate of 0.001. Adam is a variation on the standard stochastic gradient descent optimization algorithm that adjusts the learning rate based on a running average and the running variance of the recent gradients, which often speeds up convergence[13] In order to combat exploding gradients inside the LSTM layers, gradient clipping was applied with a maximum gradient norm of 1.0 and a maximum gradient value of 0.5. This technique leads to smoother training curves than does standard stochastic gradient descent. Model parameters were saved in checkpoints after each training epoch, so the parameters that resulted in the highest validation accuracies were chosen for computing additional performance metrics. In each of the roller activity tasks, both *BaselineCNN* and *DeepConvLSTM* were able to achieve very high training accuracy, but this was deemed to be overfitting, since it occurred at the expense of validation accuracy (see Figure 10). LSTMs are sometimes found to have the ability to memorize the training data, so it is not surprising that *DeepConvLSTM* achieved nearly perfect training accuracy. Its peak validation accuracy was also superior to that of *BaselineCNN*, however, so the model selected had at least some predictive value beyond mere memorization.

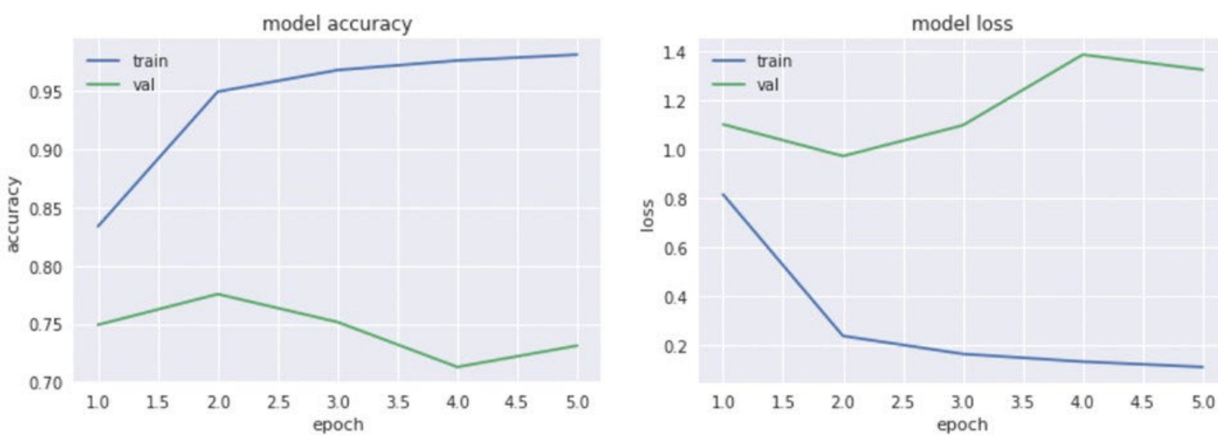

Figure 10. Accuracy and loss curves for *DeepConvLSTM* for the six-activity-class roller data

Both *BaselineCNN* and *DeepConvLSTM* were able to classify the roller's activities with reasonable accuracy, but *DeepConvLSTM*'s performance was superior. Both models showed higher performance on the easier subproblem of predicting combined classes than on the problem of predicting all six classes. As *DeepConvLSTM* was shown to be superior in identifying the roller's activities, it was the only model applied to Excavator 1.

Roller: Six-activity identification problem. For the six-activity-class problem with activities: "Forward High", "Backward High", "Forward Low", "Backward Low", "Forward Off", and "Backward Off", *BaselineCNN* had a validation accuracy of 74.2% while *DeepConvLSTM* achieved a validation accuracy of 77.1%. Table 1 summarizes precision, recall, and F1 score for both models below. Predictions for both *BaselineCNN* and

DeepConvLSTM are plotted against the true activity labels on the full data set in Figure 11. As predictions coinciding with the ground truth signal are covered by it, the number of visible peaks and troughs in the prediction signals is an indication of the degree to which they deviate from the ground truth. Furthermore, the signals are plotted with a degree of transparency, so darker lines indicate stronger signals. The predictions are not considered for the yellow shaded regions in the graphs, which were excluded from training and validation. Overall, the *DeepConvLSTM* predictions displayed in orange are a better match for the ground truth signal in both the training and the validation regions than the *BaselineCNN* predictions displayed in green.

Table 1. Roller Activity Metrics for *BaselineCNN* and *DeepConvLSTM*

Activity Label	Precision		Recall		F1-Score	
	BaselineCNN	DeepConvLSTM	BaselineCNN	DeepConvLSTM	BaselineCNN	DeepConvLSTM
Fwd. High	0.73	0.81	0.77	0.73	0.75	0.77
Bwd. High	0.81	0.75	0.34	0.32	0.47	0.45
Fwd. Low	0.65	0.72	0.67	0.8	0.66	0.76
Bwd. Low	0.76	0.75	0.91	0.93	0.83	0.83
Fwd. Off	0.87	0.80	0.72	0.9	0.79	0.85
Bwd. Off	0.69	0.86	0.99	0.86	0.82	0.86
Average	0.75	0.78	0.73	0.76	0.72	0.75

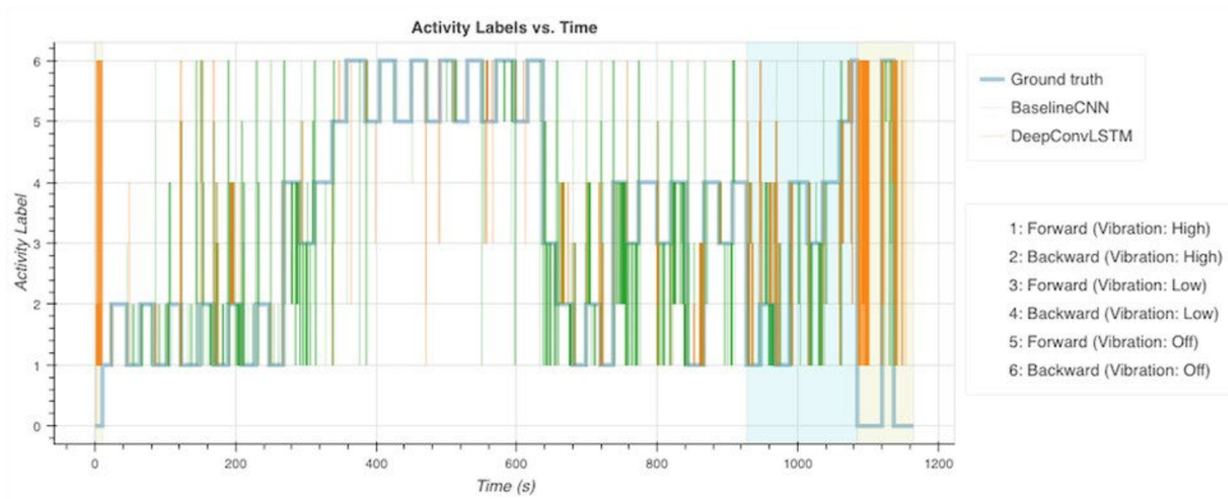

Figure 11. Roller predictions of both models compared to the ground truth data

Roller: Direction-only subproblem. In this problem, the possible activity labels were reduced to just “Forward” and “Backward”. *BaselineCNN* achieved a validation accuracy of 93.6% and an average F1 score of 0.94. *DeepConvLSTM* achieved a validation accuracy of 96.2% and an average F1 score of 0.96.

Roller: Vibration-setting only subproblem. In this problem, the possible activity labels were reduced to just the vibration settings “High,” “Low,” and “Off.” *BaselineCNN* achieved a validation accuracy of 74.4% and an average F1 score of 0.75. *DeepConvLSTM* achieved a validation accuracy of 75.2% and an average F1 score of 0.75.

Excavator 1: Seven-activity identification problem. In this problem, the possible activities were "Idling," "Traveling," "Scooping," "Dropping," "Rotating (left)," "Rotating (right)," and "Various." *DeepConvLSTM* achieved a validation accuracy of 77.6% and an average F1 score of 0.78. Although the dataset was imbalanced in favor of the "Various" activity class (over 40% of the data), counteracting the imbalance with a weighted loss function decreased the F1 score. The unweighted results were judged to be most representative and are summarized in Table 2. This preponderance of is not surprising since the "Various" label covered multiple unnamed activities, rather than being one distinct activity itself. The model struggled a little in identifying the "Traveling" activity, but this activity only comprised 2% of the data. As the confusion matrix in Figure 12(a) shows, most of the model's errors were related to the "Various" activity. To illustrate the model's predictive power beyond confusion related to the "Various" category, two additional sets of performance metrics are reported (see Table 2 and Figure 11 below). The "No-Various" and "Adjusted-Idle" results derive from an instance of *DeepConvLSTM* trained and evaluated separately on a subset of the full data set, for which every frame with the label "Various" was omitted from both training and validation (No-Various). This setup is somewhat artificial since it renders the model incapable of reasonably processing the full data set as it is. In other words, it would not know what to do with all of the "Various" labels since that category is no longer in its vocabulary. However, it provides a reasonable estimation of how the model might perform in scenarios where there is no ambiguous label like "Various". *DeepConvLSTM* managed a very high validation accuracy of 90.7%, and an average F1 score of 0.91. As the confusion matrix in Figure 12(b) suggests, the model benefitted somewhat from the fact that the removal of the "Various" activities left a disproportionately high number of "Idle" frames. Our results therefore suggests that the "Idle" activity is fairly easy to classify with extremely high accuracy. To give an estimate of the model's performance under conditions that are less favorable but still unambiguous, the same model (trained without the "Various" frames) was evaluated on a modified version of its validation set with the first 14,335 instances of "Idle" removed as well (Adjusted-Idle). Under these conditions, the class distribution in the validation data set was fairly even. *DeepConvLSTM* managed a respectable validation accuracy of 82.5% and an average F1 score of 0.83.

Table 2. Excavator 1 Activity Metrics for *DeepConvLSTM*

	Precision			Recall			F1-Score		
Activity Label	**Full data**	**No Various**	**Adjusted Idle**	**Full data**	**No Various**	**Adjusted Idle**	**Full data**	**No Various**	**Adjusted Idle**
Idling	0.90	1.00	1.00	0.97	0.96	0.81	0.93	0.98	0.89
Traveling	0.42	0.99	0.99	0.22	0.57	0.57	0.29	0.72	0.72
Scooping	0.32	0.70	0.73	0.75	0.96	0.96	0.45	0.81	0.83
Dropping	0.66	0.83	0.83	0.65	0.65	0.65	0.66	0.73	0.73
Rotating (left)	0.68	0.69	0.69	0.74	0.93	0.93	0.71	0.79	0.79
Rotating (right)	0.82	0.94	0.94	0.80	0.80	0.80	0.81	0.86	0.86
Various	0.84	N/A	N/A	0.65	N/A	N/A	0.73	N/A	N/A
Average	0.81	0.92	0.85	0.78	0.91	0.83	0.78	0.91	0.83

(a) Full data set

Predicted \ Actual	Idling	Traveling	Scooping	Dropping	Rotating (left)	Rotating (right)	Various
Idling	16731	14	130	0	305	0	110
Traveling	0	202	0	0	33	22	668
Scooping	172	0	2006	70	0	38	404
Dropping	0	0	2	1116	123	24	440
Rotating (left)	0	13	0	78	2063	172	470
Rotating (right)	0	0	152	158	92	2582	252
Various	1726	251	3917	272	401	315	12711

(b) No Various

Predicted \ Actual	Idling	Traveling	Scooping	Dropping	Rotating (left)	Rotating (right)
Idling	16566	6	694	0	24	0
Traveling	0	529	168	31	155	42
Scooping	0	0	2593	14	39	44
Dropping	0	0	15	1104	561	25
Rotating (left)	0	0	32	95	2602	67
Rotating (right)	0	0	187	81	377	2591

(c) Idle Adjusted

Predicted \ Actual	Idling	Traveling	Scooping	Dropping	Rotating (left)	Rotating (right)
Idling	2387	6	538	0	24	0
Traveling	0	529	168	31	155	42
Scooping	0	0	2593	14	39	44
Dropping	0	0	15	1104	561	25
Rotating (left)	0	0	32	95	2602	67
Rotating (right)	0	0	187	81	337	2591

Figure 12. The confusion matrices for *DeepConvLSTM*'s performance in the Excavator 1 data. Predicted labels on the vertical axis; actual labels on the horizontal axis

PHASE 2

Activity Classification. In order to compare the TCN developed with the authors' previously developed models, its performance was compared to theirs, using the same datasets. Table 3 summarizes the validation accuracies achieved on the various problems, including the dataset for Excavator 2. Wherever previous results are shown in black, new results are shown in blue. In parentheses is the number of activities in each classification task.

Table 3. Validation Accuracies of Each (Model, Experiment) Pair for the Compactor (roller) and the Excavators

	Model		
Experiment	*BaselineCNN*	*DeepConvLSTM*	**TCN**
Compactor activities (6)	74.2%	77.1%	**78.1%**
Excavator 1 activities (7)	N/A	77.6%	**8.I.4%**
Excavator 1 no Various (6)	N/A	90.70%	**91.4%**
Excavator 1 Idle adjusted (6)	N/A	82.5%	**8.I.4%**
Excavator 2 activities (7)	N/A	**6.I.41%**	78.80%

Table 4 summarizes the precision, recall, and F1-Score of each model on the vibratory roller dataset.

Table 4. Performance Metrics for Vibratory Roller Experiment

Activity Label	Precision			Recall			FI -Score		
	BaselineCNN	DeepConv-LSTM	TCN	BaselineCNN	DeepConv-LSTM	TCN	BaselineCNN	DeepConv-LSTM	TCN
Fwd. High	0.73	0.81	0.78	0.77	0.73	1.79	0.75	0.77	0.79
Bwd. High	0.81	0.75	0.89	0.34	0.32	0.30	0.47	0.45	0.44
Fwd. Low	0.65	0.72	0.73	0.67	0.8	1.74	0.66	0.76	0.73
Bwd. Low	0.76	0.75	0.75	0.91	0.93	1.95	0.83	0.83	0.84
Fwd. Off	0.87	0.8	0.88	0.72	0.9	1.87	0.79	0.85	0.88
Bwd. Off	0.69	0.86	0.8	0.99	0.86	1.97	0.82	0.86	0.88
Average	0.75	0.78	0.79	0.73	0.76	1.78	0.72	0.75	0.76

Tables 5–7 summarize the precision and recall of each model in the experiments involving Excavator 1. The confusion matrices for *DeepConvLSTM* and TCN in the Excavator 2 experiments are shown in Figure 13 and 14, respectively.

Table 5. Performance Metrics for "Full" Excavator 1 Experiment

Activity Label	Full data					
	Precision		Recall		FI-Score	
	DeepConvLSTM	TCN	DeepConvLSTM	TCN	DeepConvLSTM	TCN
Idling	0.90	0.95	0.97	0.98	0.93	0.97
Traveling	0.42	0.51	0.22	0.53	0.29	0.52
Scooping	0.32	0.61	0.75	0.15	0.45	0.24
Dropping	0.66	0.85	0.65	0.43	0.66	0.57
Rotating (left)	0.68	0.65	0.74	0.74	0.71	0.69
Rotating (right)	0.82	0.72	0.80	0.82	0.81	0.77
Various	0.84	0.80	0.65	0.87	0.73	0.83
Average	0.81	0.83	0.78	0.83	0.78	0.82

Table 6. Performance Metrics for "No-Various" Excavator 1 Experiment

Activity Label	No Various					
	Precision		Recall		FI-Score	
	DeepConvLSTM	TCN	DeepConvLSTM	TCN	DeepConvLSTM	TCN
Idling	1.00	0.98	0.96	0.99	0.98	0.99
Traveling	0.99	0.91	0.57	0.54	0.72	0.68
Scooping	0.70	0.8l	0.96	0.79	0.81	0.81
Dropping	0.83	0.78	0.65	0.62	0.73	0.69
Rotating (left)	0.69	0.79	0.93	0.83	0.79	0.81
Rotating (right)	0.94	0.80	0.80	0.92	0.86	0.88
Various	N/A	N/A	N/A	N/A	N/A	N/A
Average	0.92	0.91	0.91	0.91	0.91	0.91

Table 7. Performance Metrics for "Adjusted-Idle" Excavator 1 Experiment

Activity Label	Adjusted Idle					
	Precision		Recall		FI-Score	
	DeepConvLSTM	TCN	DeepConvLSTM	*TCN*	*DeepConvLSTM*	TCN
Idling	1.00	0.90	0.81	1.00	0.89	0.94
Traveling	0.99	0.91	0.57	0.54	0.72	**0.68**
Scooping	0.73	0.86	0.96	0.79	0.83	0.82
Dropping	0.83	0.78	0.65	0.62	0.73	0.69
Rotating (left)	0.69	0.79	0.93	0.8l	0.79	0.81
Rotating (right)	0.94	0.89	0.80	0.92	0.86	**0.88**
Various	N/A	N/A	N/A	N/A	N/A	N/A
Average	0.85	0.84	0.83	0.8l	0.83	0.83

Figure 13. Confusion matrix for *DeepConvLSTM* in the Excavator 2 experiment

Figure 14. Confusion matrix for TCN in the Excavator 2 experiment

Linking Activities to Emissions. The Freedman-Diaconis rule was used to select appropriate bin sizes for the histograms as it makes no assumptions about the distribution it is modeling while attempting to minimize the difference between the empirically-derived histogram and the theoretical probability distribution. The resulting histograms are plotted in Figures 15–17, superimposed on "kernel density estimations" of the true distributions.

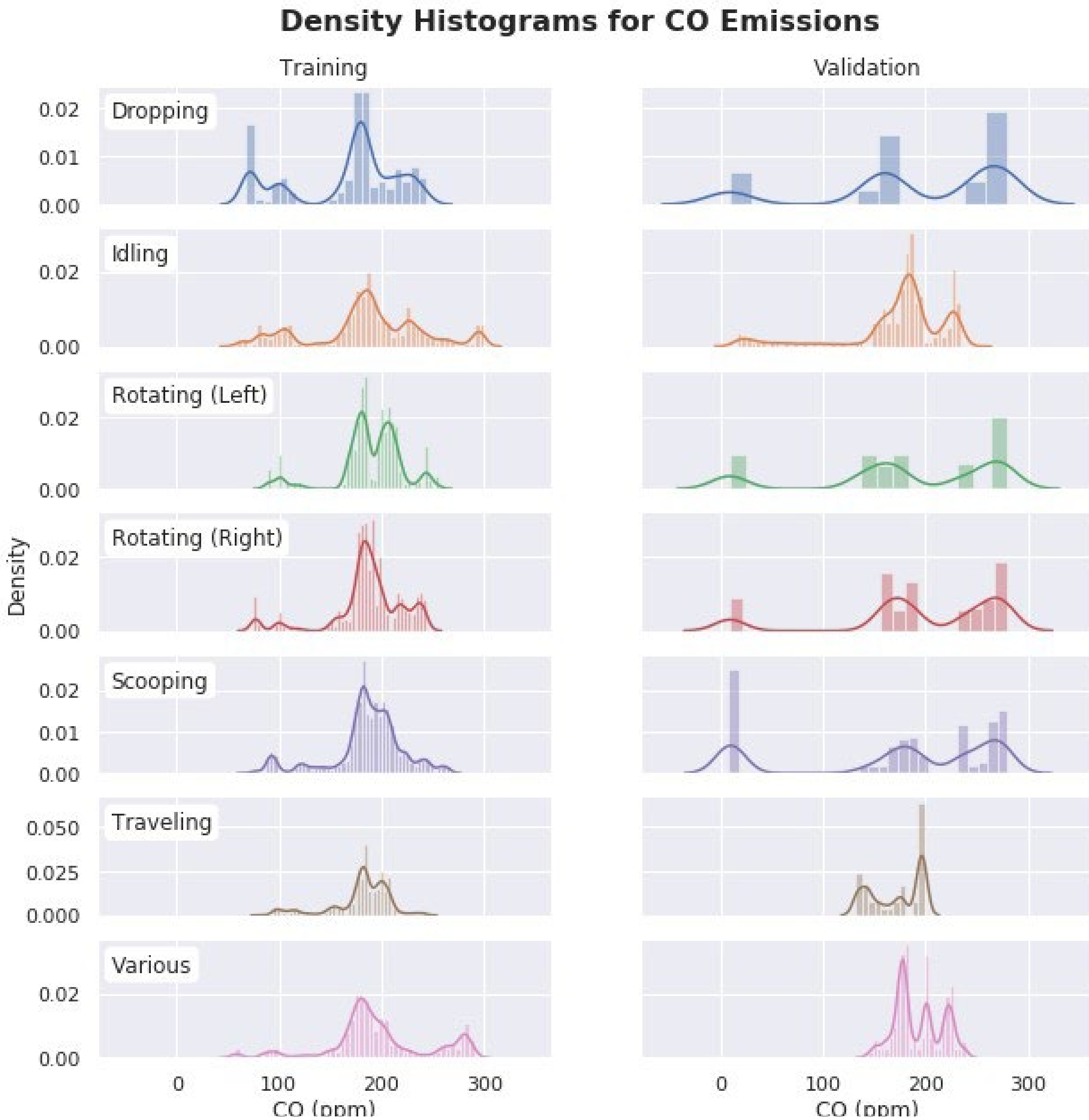

Figure 15. Density histograms for CO emissions across training and validation sets

Figure 16. Density histograms for NOx emissions across training and validation sets

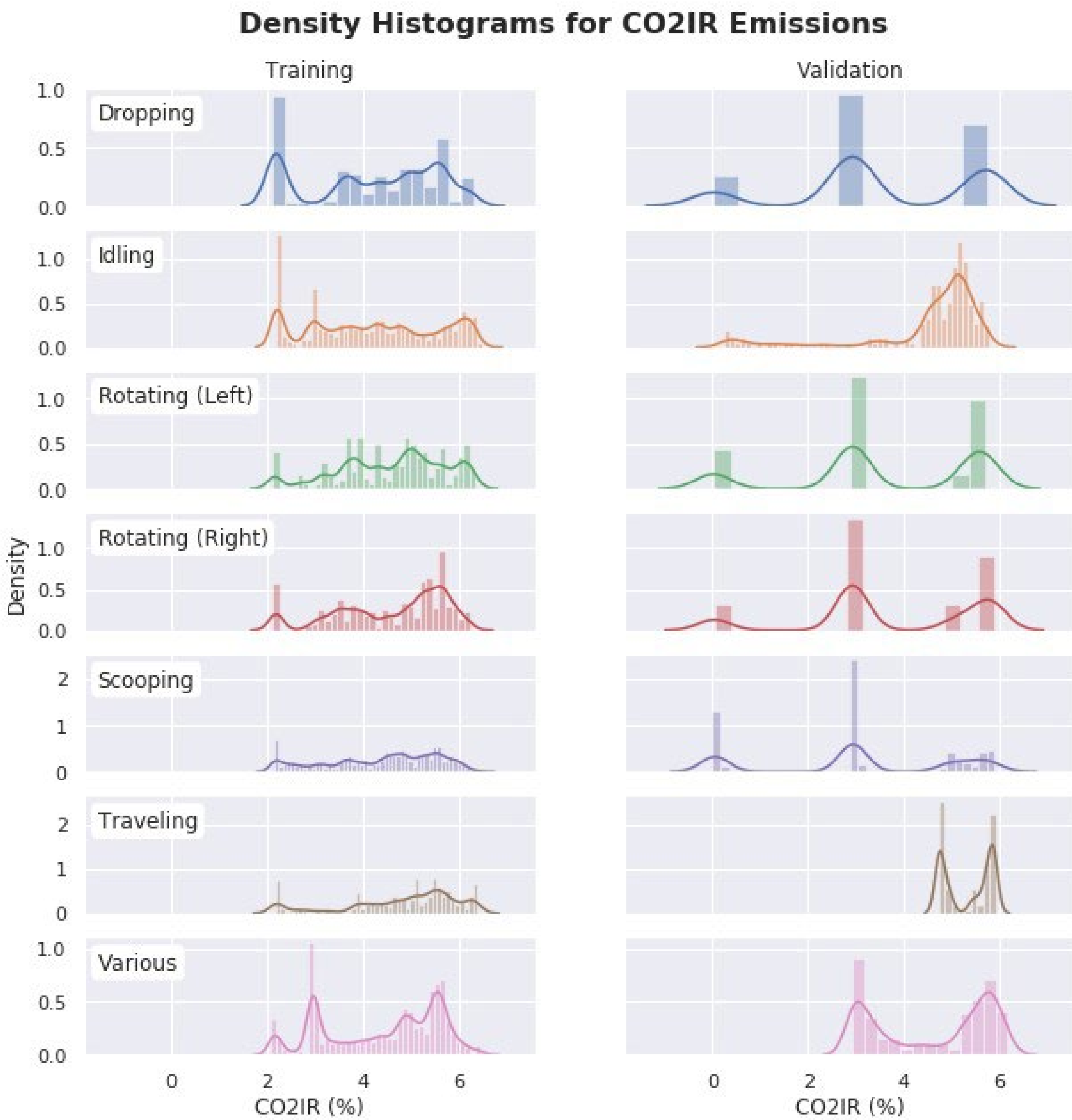

Figure 17. Density histograms for CO2 emissions across training and validation sets

V. DISCUSSION AND CONCLUSION

It was observed across all of the measurements taken that the new TCN model is at least competitive with the previous reigning champion, *DeepConvLSTM*. In fact, it beats *DeepConvLSTM* in terms of validation accuracy every time, despite training much faster and being simpler to explain. The most notable differences in performance occurred in the two excavator experiments, which are challenging datasets because they include many activities that sometimes occur together to such a degree that the authors could only label the activity during such instances as *Various.* In the first excavator experiment, *DeepConvLSTM* managed a validation accuracy of 77.6%, with its mistakes largely coming from confusion related to the *Various* label. Eliminating the *Various* label from consideration and rebalancing the data set by adjusting the remaining number of *Idle* labels allowed its accuracy to rise to 82.5%, but the TCN managed to achieve 83.4% validation accuracy regardless of whether the *Various* label was present. In the second excavator experiment, we see a similar trend (see the confusion matrices in Figures 13 and 14 above); again it seems that *DeepConvLSTM* tended to get confused by the *Various* label, but the TCN faired much better. The TCN achieved a validation accuracy of 78.8%—this time much higher than the 63.4% managed by *DeepConvLSTM* (see Table 3).

Both *DeepConvLSTM* and TCN displayed distinct noteworthy trends during training. In general, *DeepConvLSTM* manages to train to nearly perfect training accuracy (95%+) on every data set studied, but it fails to generalize well from this fitting (as judged from its validation accuracy), despite numerous attempts at different forms of regularization. In contrast, the TCN model tends to have a training accuracy very similar to its validation accuracy, suggesting that it tends to overfit less often than *DeepConvLSTM*, but also that it might have lower capacity to express very complicated patterns in the training data. It might be possible to rule out this tendency to overfit and take better advantage of this expressiveness with a larger data set—larger not just in terms of samples, but expanded in real-world terms like the number of machines, the number of activities, the time spent performing them, performing work on different days, etc—and in such a scenario, *DeepConvLSTM* still seems promising.

ENDNOTES

1. Hajji, Apif M., and Phil Lewis. “Development of productivity-based estimating tool for energy and air emissions from earthwork construction activities.” Smart and Sustainable Built Environment 2, no. 1 (2013): 84–100. doi: 10.6106/JCEPM.2013.3.2.058

2. Kenley, R., and T. Harfield. “Greening procurement of infrastructure construction: Optimizing mass haul operation to reduce greenhouse gas emissions.” In *Proceedings of the CIB W78–W102 International Conference, Sophia Antipolis, France*, pp. 26–28. 2011.

3. Vorster, Michael C., and Jesus M. De La Garza. “Consequential equipment costs associated with lack of availability and downtime.” *Journal of Construction Engineering and Management* 116, no. 4 (1990): 656–669. doi: 10.1061/(ASCE)0733-9364(1990)116:4(656)

4. McCahill, Dennis F., and Leonhard E. Bernold. “Resource-oriented modeling and simulation in construction.” *Journal of Construction Engineering and Management* 119, no. 3 (1993): 590–606. doi: 10.1061/(ASCE)0733-9364(1993)119:3(590)

5. Environmental Protection Agency Clean Air Act Advisory Committee. “Recommendations for reducing emissions from the Legacy Diesel Fleet.” *US Environmental Protection Agency, Washington DC* (2006).

6. Ahn, Changbum R., SangHyun Lee, and Feniosky Peña-Mora. “Monitoring system for operational efficiency and environmental performance of construction operations using vibration signal analysis.” In *Construction Research Congress 2012: Construction Challenges in a Flat World*, pp. 1879–1888. 2012. doi: 10.1061/9780784412329.189

7. Akhavian, Reza, and Amir H. Behzadan. “Simulation-based evaluation of fuel consumption in heavy construction projects by monitoring equipment idle times.” In *2013 Winter Simulations Conference (WSC)*, pp. 3098–3108. IEEE, 2013. doi: 10.1109/WSC.2013.6721677

8. Ahn, Changbum R., SangHyun Lee, and Feniosky Peña-Mora. “Application of low-cost accelerometers for measuring the operational efficiency of a construction equipment fleet.” *Journal of Computing in Civil Engineering* 29, no. 2 (2013): 04014042. doi: 10.1061/(ASCE)CP.1943-5487.0000337

9. Frey, H. Christopher, William Rasdorf, and Phil Lewis. “Comprehensive field study of fuel use and emissions of nonroad diesel construction equipment.” *Transportation Research Record* 2158, no. 1 (2010): 69–76. doi: 10.3141/2158-09

10. Akhavian, Reza, and Amir H. Behzadan. “Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers.” Advanced Engineering Informatics 29, no. 4 (2015): 867–877. doi: 10.1016/j.aei.2015.03.001

11. Ordóñez, Francisco, and Daniel Roggen. “Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition.” Sensors 16, no. 1 (2016): 115. doi: 10.3390/s16010115

12. Van den Oord, A., S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu, WaveNet: A Generative Model for Raw Audio, ArXiv:1609.03499 [Cs]. (2016). http://arxiv.org/abs/1609.03499 (accessed July 6, 2019).

13. Kungma, D P, and J Ba. 2014. “Adam: A Method for Stochastic Optimization.” Accessed March 7, 2019. http://arxiv.org/abs/1412.6980

BIBLIOGRAPHY

Ahn, Changbum R., SangHyun Lee, and Feniosky Peña-Mora. “Application of low-cost accelerometers for measuring the operational efficiency of a construction equipment fleet.” Journal of Computing in Civil Engineering 29, no. 2 (2013): 04014042. doi: 10.1061/(ASCE)CP.1943-5487.0000337

Ahn, Changbum R., SangHyun Lee, and Feniosky Peña-Mora. “Monitoring system for operational efficiency and environmental performance of construction operations using vibration signal analysis.” In Construction Research Congress 2012: Construction Challenges in a Flat World, pp. 1879–1888. 2012. doi: 10.1061/9780784412329.189

Akhavian, Reza, and Amir H. Behzadan. “Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers.” Advanced Engineering Informatics 29, no. 4 (2015): 867–877. doi: 10.1016/j.aei.2015.03.001

Akhavian, Reza, and Amir H. Behzadan. “Simulation-based evaluation of fuel consumption in heavy construction projects by monitoring equipment idle times.” In 2013 Winter Simulations Conference (WSC), pp. 3098–3108. IEEE, 2013. doi: 10.1109/WSC.2013.6721677

Environmental Protection Agency Clean Air Act Advisory Committee. “Recommendations for reducing emissions from the Legacy Diesel Fleet.” US Environ-mental Protection Agency, Washington DC (2006).

Frey, H. Christopher, William Rasdorf, and Phil Lewis. “Comprehensive field study of fuel use and emissions of nonroad diesel construction equipment.” Transportation Research Record 2158, no. 1 (2010): 69–76. doi: 10.3141/2158-09

Hajji, Apif M., and Phil Lewis. “Development of productivity-based estimating tool for energy and air emissions from earthwork construction activities.” Smart and Sustainable Built Environment 2, no. 1 (2013): 84–100. doi: 10.6106/JCEPM.2013.3.2.058

Kungma, D P, and J Ba. 2014. “Adam: A Method for Stochastic Optimization.” Accessed March 7, 2019. http://arxiv.org/abs/1412.6980

Kenley, R., and T. Harfield. “Greening procurement of infrastructure construction: Optimizing mass haul operation to reduce greenhouse gas emissions.” In Proceedings of the CIB W78–W102 International Conference, Sophia Antipolis, France, pp. 26–28. 2011.

McCahill, Dennis F., and Leonhard E. Bernold. “Resource-oriented modeling and simulation in construction.” Journal of Construction Engineering and Management 119, no. 3 (1993): 590–606. doi: 10.1061/(ASCE)0733-9364(1993)119:3(590)

Ordóñez, Francisco, and Daniel Roggen. “Deep convolutional and lstm recur-rent neural networks for multimodal wearable activity recognition.” Sensors 16, no. 1 (2016): 115.

Van den Oord, A., S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu, WaveNet: A Generative Model for Raw Audio, ArXiv:1609.03499 [Cs]. (2016). http://arxiv.org/abs/1609.03499 (accessed July 6, 2019).

Vorster, Michael C., and Jesus M. De La Garza. “Consequential equipment costs associated with lack of availability and downtime.” Journal of Construction Engineering and Management 116, no. 4 (1990): 656–669. doi: 10.1061/(ASCE)0733-9364(1990)116:4(656)

ABOUT THE AUTHOR

REZA AKHAVIAN, PHD

Dr. Reza Akhavian is an Assistant Professor at the Department of Civil, Construction, and Environmental Engineering at San Diego State University (SDSU) and the Director of the Data-informed Construction Engineering (DiCE) Group. Before that, he was an Assistant Professor of Construction Management at the School of Engineering, California State University East Bay (CSUEB). He received his Ph.D. in Civil Engineering (with Construction Engineering and Management concentration) from the University of Central Florida (UCF). He also holds an M.S. (University of Central Florida, 2012) and a B.S. (University of Tehran, 2010) in Civil Engineering. He has more than 40 articles published in peer-reviewed journals and conference proceedings and serves as a member of the editorial board of the American Society of Civil Engineers (ASCE) Journal of Construction Engineering and Management (JCEM) and Elsevier Journal of Advanced Engineering Informatics. His industry experience includes working with Bentley Systems and DPR Construction and advising construction technology startups. Dr. Akhavian's research interests are in the areas of information technology, data analytics, machine learning, and automation and robotics for different civil engineering and infrastructure management applications. More specifically, his research and industry involvement focus on leveraging different types of data collected from jobsite entities such as workers, equipment, and material and the built environment to help make informed-decisions. Much current work involves enhancing health, safety, and productivity of construction project operations and sustainability and resiliency of the built environment.

PEER REVIEW

San José State University, of the California State University system, and the Mineta Transportation Institute (MTI) Board of Trustees have agreed upon a peer review process required for all research published by MTI. The purpose of the review process is to ensure that the results presented are based upon a professionally acceptable research protocol.