Accepted manuscript
doi: 10.1680/jsmic.21.00031

Accepted manuscript

As a service to our authors and readers, we are putting peer-reviewed accepted manuscripts (AM) online, in the Ahead of Print section of each journal web page, shortly after acceptance.

Disclaimer

The AM is yet to be copyedited and formatted in journal house style but can still be read and referenced by quoting its unique reference number, the digital object identifier (DOI). Once the AM has been typeset, an 'uncorrected proof' PDF will replace the 'accepted manuscript' PDF. These formatted articles may still be corrected by the authors. During the Production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal relate to these versions also.

Version of record

The final edited article will be published in PDF and HTML and will contain all author corrections and is considered the version of record. Authors wishing to reference an article published Ahead of Print should quote its DOI. When an issue becomes available, queuing Ahead of Print articles will move to that issue's Table of Contents. When the article is published in a journal issue, the full reference should be cited in addition to the DOI.

Submitted: 15 October 2021

Published online in 'accepted manuscript' format: 02 April 2022

Manuscript title: Automating excavator productivity measurement using deep learning

Authors: Elham Mahamedi[1], Kay Rogage[2], Omar Doukari[1], Mohamad Kassem[1]

Affiliations: [1]Department of Mechanical & Construction Engineering, Northumbria University, Newcastle-Upon-Tyne, UK. [2]Department of Computer and Information Sciences, Northumbria University, Newcastle-Upon-Tyne, UK.

Corresponding author: Mohamad Kassem, Department of Mechanical & Construction Engineering, Northumbria University, Newcastle-Upon-Tyne, UK.

E-mail: mohamad.kassem@northumbria.ac.uk

Abstract

Heavy equipment represents a major cost element and a critical resource in large infrastructure projects. Automating the measurement of their productivity is important to remove the inaccuracies and inefficiencies of current manual measurement processes and to improve the performance of projects. Existing studies have prevalently focused on equipment activity recognition using mainly vision based systems which require intrusive field installation and the application of more computationally demanding methods. This study aims to automate the measurement of equipment productivity using a combination of smartphone sensors to collect kinematic and noise data and deep learning algorithms. Different combination inputs and deep learning methods were implemented and tested in a real-world case study of a demolition activity. The results demonstrated very high accuracy (99.78%) in measuring the productivity of the excavator. Construction projects can benefit from the proposed method to automate productivity measurement, identify equipment inefficiencies in near real-time, and inform corrective actions.

doi: 10.1680/jsmic.21.00031

Notation

Z	is the downtime ratio
D	is the number of hours a particular equipment unit is broken down in a month
W	is the total number of hours worked by the equipment in the month
DT	is equipment downtime
UR	is the utilization ratio

Accepted manuscript
doi: 10.1680/jsmic.21.00031

1. Introduction

Equipment productivity is critical to the success of construction projects, particularly in equipment-intensive projects such as earth-moving, pavement and tunnel projects. Construction equipment productivity has been studied by many researchers (Ok and Sinha, 2006; Gurmu and Aibinu, 2017; Gerami Seresht and Fayek, 2018) to improve the overall construction productivity and reduce project time and cost. To improve productivity, it must be measured and monitored throughout the project execution phase to identify equipment inefficiencies and their root causes. However, collecting the required data for equipment performance monitoring is time and resource consuming (Chen et al., 2020). Manual data collection particularly, is error prone and impracticable in large projects (Kim et al., 2018). This demonstrates the need for automating the process of data collection about equipment operation, measuring and analysing their productivity, and monitoring their performance in large construction projects.

Data is key to enable these capabilities. Three main approaches for data collection about equipment exist. The first approach is enabled by Original Equipment Manufacturer (OEM) through on-board integrated telematics. However, a review of such systems reveals inconsistencies among OEMs in terms of type of data collected, their definition (e.g. idle time) and reporting intervals (Jagushte, 2017; Kassem et al., 2019). These inconsistencies affect the usability of telematics data especially in the case of mixed equipment fleet. An alternative approach is to use vision-based systems (Chen et al. 2020)). However, vision-based approaches focus mainly on equipment activity recognition, do not offer a productivity measurement

doi: 10.1680/jsmic.21.00031

approach, require a laborious field installation, and entails computationally demanding methods. A third approach is to use sensors such as smartphone phone sensors and apps (e.g. gyroscope, accelerometer, noise). Compared to the vision-based approach, this approach requires less intrusive installation and lower computational resources. However, there are only a few studies analysing its performance in terms of measuring the productivity of site equipment. The aim of this paper is to automate the measurement of equipment productivity by combining smartphone sensors and deep learning techniques in order to collect the required data, automate feature extraction in complex activity recognition, and perform the productivity measurement.

This remainder of the paper is structured as follows. Section 2 summarises existing studies on automation of equipment activity recognition and productivity measurement. Section 3 introduces some of the key equipment productivity metrics. Section 4 explains the proposed approach by this study. Section 5 demonstrates its testing in a real case study. Section 6 discusses the findings and states the limitations. Section 7 presents the limitations and future work. Finally, Section 7 Concludes.

2. Related Studies

The traditional approaches for measuring equipment productivity are time consuming and error prone as they are based on manual data collection, direct observation of activities using sampling or survey (Akhavian and Behzadan, 2015). Therefore, automating equipment productivity measurement is an essential need to monitor and enhance equipment performance, particularly in large-scale projects. Several studies have been carried out to recognize equipment activities, determine their activity duration, and identify their operation cycle time through automated data capture. They used different technologies for data collection including sensors, computer vision and audio signals. These approaches are reviewed in this section.

Montaser and Moselhi (2012) proposed an approach for tracking earthmoving operations using Radio Frequency Identification (RFID). Their approach could automatically recognize four states of the trucks including loading, travelling, dumping and returning. As this approach uses fixed RFID readers for gate systems at the loading and dumping areas, it would be more relevant to projects with fixed loading and dumping areas. Moreover, this approach cannot identify the waiting time of the trucks in the loading/dumping areas. In another study, Montaser and Moselhi (2014) developed an automated system integrating Global Positioning System (GPS) and Geographical Information System (GIS). This system tracks the location of the trucks using GPS units mounted on the trucks and identifies the spatial boundaries of loading and dumping areas using GIS. Similar to their previous approach, they recognized the same four states for the trucks, but this system also lacked the capability of capturing waiting times in the loading/dumping areas. To address this drawback and to improve accuracy of measuring excavated soil volume, Ibrahim and Moselhi (2014) developed an automated productivity assessment method for earthmoving operations. They adopted mobile sensors including GPS mounted on trucks to track their locations, accelerometers mounted on the bed of the trucks for tilt sensing of the truck bed, strain gauges mounted on truck leaf springs to measure soil weight, barometric pressure sensors attached to the bucket of loaders to measure elevation of the buckets, and Bluetooth-based RF module for data transfer and proximity detection between equipment. They developed an algorithm to use the collected data from these sensors for the truck activity recognition including load queue, load, travel, dump queue, dump, return and service. The developed method measured productivity with only 2.2% error. Despite its high accuracy and the simplicity of their data processing algorithm, the implementation of this method requires extensive installation of several sensors on the trucks and loaders, which are not often possible in construction projects due to accessibility and availability issues and the ownership models of heavy equipment.

Ahn et al. (2012) utilised an accelerometer mounted inside the cabin of a medium-sized excavator collecting the data with a frequency of 100 Hz. They presented the relationship between operational efficiency and environmental performance using vibration signals. A further study by Ahn et al. (2015) explored capturing acceleration signals from four types of excavators using an accelerometer mounted inside the cabin and conducted the experiment under an instructed environment. The experiment involved the operation of an excavator that was strictly instructed to capture the required data in order to analyse patterns of accelerometer data. They used different supervised classifiers including Naïve Bayes, Instance-based learning, K-nearest neighbour (KNN) and Decision tree (J48) and achieved over 93% accuracy for classification of excavators' operation.

One study explored approaches to detecting loading and unloading of a dumper truck with a remote tracking technique using 3-axis magnetic field sensing and 3-axis tilt sensing for a loader and a truck in an indoor laboratory (Akhavian and Behzadan, 2012). Akhavian and Behzadan (2015) also developed an automated method to detect equipment activities and their durations for simulation input modelling of a front-end loader using GPS sensor, 3-axis accelerometer, and 3-axis gyroscope with frequency of 100 Hz. This technique applied several supervised learning methods including logistic regression, K-NN, decision tree, neural network, support vector machine (SVM), and achieved an overall accuracy of 86%.

Some studies used Inertial Measurement Unit (IMU) data from the sensors embedded in smartphones including accelerometers and gyroscopes for equipment activity recognition. For instance, Kim et al. (2018) measured an excavator operation cycle time using IMU data with the frequency of 128 Hz. They applied Random Forest, Naïve Bayes, J48 and Sequential Minimal Optimization (SMO) for the cycle time prediction and achieved 91.83% accuracy. In another study Rashid and Louis (2019) used time-series data augmentation on 3-axis accelerometer, and 3-axis gyroscope data collected with the frequency of 80 Hz to generate synthetic training data for four types of excavators and front-end loaders. This technique applied recurrent neural network (RNN) and achieved over 96% accuracy for fourfold augmentation.

Kassem et al. (2021) developed a DNN model for measuring the volume of earth excavated from a mixed fleet of excavators (e.g. different sizes, weights, models) and benchmarked the performances of excavation work using telematics data from 21 days of operation. They achieved an accuracy of 69.64% which was deemed acceptable due to the involvement of the manual work (i.e. archaeologists) alongside the equipment in the selected central London case study.

Bae et al. (2019) developed a dynamic time warping algorithm for activity identification and automatic classification of excavator activities (i.e., digging, levelling, lifting, trenching, traveling, and idling) using joysticks signals. The correct-recognition rate of their model was between 91% and 97%.

Despite the contributions these studies bring to monitoring construction equipment activity, very few studies have attempted to automate equipment productivity measurement. One recent study by Chen et al. (2020), developed a vision-based method for measuring excavator productivity. However, this method revealed computationally expensive and had some limitations such as dependency of the results on the light conditions, viewpoints of cameras, number of equipment in the scene and background movements. In addition, their achieved accuracy was 83% for productivity measurement, and 94% for idle time measurement.

In (Bügler et al., 2017), photogrammetry was combined with video analysis for measuring the volume of the excavated soil and computing soil removal productivity. Kim et al. (2019) developed a model adopting computer vision and simulation for the analysis of earthmoving productivity. They used videos from surveillance cameras at the entrance and exit of a construction site for license plate detection of dump trucks, which could produce the site access log, then analysed the truck productivity through a simulation model. Torres Calderon et al., (2021) could improve the capabilities of vision-based activity analysis methods by developing a new approach using the data synthesized from 3D kinematically configurable models for training the computer vision algorithms.

The main advantage of using computer vision for equipment productivity is that visual

doi: 10.1680/jsmic.21.00031

data can provide information about the physical movements of equipment and their visual features and spatial contextual natures (Kim and Chi, 2020). However, from practicality perspectives, this method has some limitations and challenges including:

- sensitivity to environmental factors such as occlusions, lighting, and illumination conditions (Cheng et al., 2017);
- Shaking of cameras caused by wind, and blur of images caused by rain, snow, and fog (Gong and Caldas, 2011); and
- The need for installing multiple cameras for covering a large job (Cheng et al., 2017).

Audio has been another source of data for equipment activity recognition as heavy equipment generally generates distinct acoustic patterns while performing routine tasks (Cheng et al., 2019). Cheng et al. (2017) classified the equipment activities into two states, productive or major activities and non-productive or minor activity, to recognise the equipment states using the sound generated by construction equipment. (Sabillon et al., 2020) developed a model to use audio data for estimating the cycle time of equipment.

Audio signals are easier to use comparing to sensors and computer vision methods because their capturing technologies such as microphone, can cover a large area. In addition, processing audio files is computationally less expensive (Sabillon et al., 2020). However, background noise can impact the accuracy of the models, and some equipment does not generate distinct sound patterns during operation, which makes it difficult to recognise their activity (Cheng et al., 2017).

This study contributes to this research domain by developing and testing a new method based on smartphone sensors and deep learning to predict equipment productivity with high accuracy through a low cost and easy-to-install system on equipment.

3. Equipment productivity metrics

Productivity is generally defined as the ratio of output over input. Different metrics have been proposed to measure equipment productivity and evaluate efficiency of equipment usage. For instance, some metrics have accounted for downtime for evaluating equipment productivity. Vorster and De La Garza (1990) defined the downtime ratio (Z) for equipment over a month, as shown in Equation 1:

$$Z(\%) = \frac{D}{D+W} \times 100 \tag{1}$$

where D is the number of hours a particular equipment unit is broken down in a month, and W is the total number of hours worked by the equipment in the month.

Nepal and Park (2004) defined equipment downtime (DT) ratio as shown in Equation 2:

$$DT(\%) = \frac{\text{Total DT hours}}{\text{Total planned working hours}} \times 100 \tag{2}$$

Utilisation Ratio (UR) is another productivity metric which accounts for the percentage of time that an equipment is available for operation. It is expressed as the ratio between the total working time of an equipment and the total time available for an equipment (e.g. 24 hours or shift time) as expressed in Equation (3) (Ibbs Jr and Terveer, 1984):

$$UR(\%) = \frac{\text{Total working time}}{\text{Total available time}} \times 100 \tag{3}$$

if the data to support these metrics can be automatically collected and deep learning is used to produce reliable predictions, these metrics can be used to measure the productivity of different types of equipment, hence; enabling monitoring and benchmarking of their performance and identification of underperforming equipment.

4. Methodology

As shown in Section 3, a range of metrics for measuring equipment productivity exist. This paper adopts the utilization ratio metric (Equation 3) for measuring productivity by identifying active and non-active states of an equipment during its available time for use (e.g. day shift). Active state relates to the time that the equipment is actively working. Inactive state relates to the time that the equipment is not working including the idle time and the time the equipment engine is off. Figure 1 illustrates how these data are used to identify equipment states.

The first step is to capture data. Built-in smartphones sensors are used to capture IMU data (i.e., tri-axial accelerometer, gyroscope and linear acceleration data) and noise level data from inside the equipment operator cabins. Videos to identify when the equipment is active or inactive are captured using a camera and are used for labelling the time-stamped sensor data and developing and validating machine learning models.

The second step is data preprocessing, in which sliding windows to divide input signal data into windows of signals are identified. Each window has a few seconds of observation data. The size of each sliding window, which depends on the model specifications such as the data type and nature of the activities to be classified, affects the model size and training speed: the smaller the window size, the smaller the model and the faster the training speed (Banos et al., 2014). That is, reducing the window size enables faster activity recognition and less computational burdens. Large windows are generally used for identifying complex activities

(Banos et al., 2014). After selecting a suitable sliding window size, the data is labelled with the equipment states (i.e., either active or inactive) using the observations from the captured videos. One window would represent one sample. In this study 5 seconds of data was chosen for the window size as it had the appropriate fit to the model in terms of performance and prevented overfitting. Some applications tolerate having adjacent windows overlap, but this approach is less commonly used (Banos et al., 2014). Additionally, the overlap window can lead to an increased training dataset and overfitting. As the frequency of data capturing was 8 Hz, one sample data would represent 40 sensor datapoints. A label was generated for each window by finding the most frequently used label among the set of datapoints within each window.

The pre-processed data are then fed to the deep learning model for classification. Deep learning algorithms are more suitable for complex activity recognition because they automate feature engineering and extraction (as one of the most important and challenging tasks in machine learning) and extract high-level representation in deep layers (Wang et al., 2019).

In this study three deep learning algorithms, namely, Deep Neural Network (DNN), Convolutional Neural Network-Long Short-Term memory network (CNN-LSTM) (Mutegeki and Han, 2020, Donahue et al., 2015) and Convolutional Long Short-Term Memory (Conv-LSTM) (Xingjian et al., 2015) were tested. These algorithms are commonly used for activity recognition due to their deep structures for automated feature extractions from raw sensor data with random noises (Mahmud et al., 2020). These algorithms are applied to a various combination of data collected in a case study to compare their performance in predicting equipment states and measuring equipment productivity. The description of these algorithms and their configuration for this study is summarized below.

4.1 Deep neural network (DNN)

DNN is an extension of the multilayer perceptron (MLP) neural network, that has more than one hidden layer. DNN map inputs to outputs through a sequence of data transformations (layers). In the learning process of DNN, the values of the parameters (weights) of the layers are identified in such a way that the network correctly maps the input data to output data (i.e., minimizing the error) (Chollet, 2017). DNN is computationally complex because many parameters exist for each layer and a change in one parameter will impact other parameter behaviours (Chollet, 2017). More (deep) layers in DNN comparing to the traditional neural networks, make it more suitable for building learning models from a large amount of data, where manually extracting features is too complex or time consuming for building a successful model.

In DNN, different types of layers such as dense, flatten, dropout and sigmoid and softmax functions can be used. Dense layers are a regular neuron layer, which are densely connected and receive input from the previous layer and send output to the next layer. The input and output are also connected by the weights. Flatten layers are used to make multidimensional output linear to pass it to the dense layer when required. Dropout is a regularization method, which randomly (at a probability) drops some neurons to prevent overfitting the model. Sigmoid or softmax can be used before the output layer to output a probability distribution over the different output classes, which identifies the probability that the sample belongs to a specific class (Chollet, 2017). Figure 2 (a) depicts the architecture of the DNN model created in this study. It shows that it consists of three hidden dense layers or fully connected layers, which was found in our experiments sufficient to make the model deep enough to achieve the high-accuracy results. The detail of the model architecture is presented in Figure 3. In this model, Rectified Linear Unit (ReLU) is used as the activation function for the hidden layers which is a widely used activation function in deep learning. ReLU will return the input value if it is positive, and it will return 0 if the input value is negative. For the output layer, the sigmoid activation function is used. Sigmoid is mostly used for the binary classification and maps the input into a value ranging from 0 to 1.

4.2 Convolutional Neural Network-Long Short-Term memory network (CNN-LSTM)

In the CNN-LSTM method, Convolutional Neural Network (CNN) is created and followed by long short-term memory network (LSTM), and a dense layer on the output.

CNNs are characterized by the ability of easy training, knowledge extraction and feature extraction on input data (Huang and Kuo, 2018). CNNs are mostly adopted for image processing. LSTM is a type of Recurrent neural networks (RNNs), which are used to learn from sequence data (i.e., sequences of observations over time) and can address some difficulties of RNN in training a stable model (Brownlee, 2016). LSTM develops internal representation of the input while reading input observations in sequence and focusing on model prediction errors in the input sequence in each defined window, which is called backpropagation over time (Brownlee, 2016).

In the CNN-LSMT architecture, 1) CNN performs feature extraction on input data through convolutional layers (e.g. Conv1D), which performs convolution operations to learn local patterns (while dense layers learn global patterns) (Chollet, 2017), and pooling layers, which performs a down sampling operation to produce the most significant features (Swapna et al., 2018), and LSTM supports sequence prediction, 2) data are read sequentially in blocks and features are extracted from each block, and 3) the extracted features are fed into LSTM for interpretations and predictions (Brownlee, 2018). CNN-LSTM is more efficient for recognition of activities with differing time spans such as visual time series prediction problems. As CNN is a specific type of DNN, DNN layers can also be used in CNN-LSTM models. The CNN-LSTM model used in this study is illustrated in Figure 2 (b) and detailed in Figure 4. The proposed CNN-LSTM model consists of two consecutive blocks of 1D convolutional layer with 64 filters. The rectified linear function (ReLU) is used as the activation function. After the second 1D-CNN layer is processed, the dropout technique is applied. Next, a pooling layer is applied to reduce the number of parameters and computation in the network and to avoid overfitting. A LSTM layer with 128 units is then applied before a dropout layer for extracting temporal features. The LSTM layer is followed by a dense layer with 128 neurons and ReLU activation function. At the final step, another dense layer with a sigmoid function is added.

4.3 Convolutional Long Short-Term Memory (Conv-LSTM)

Conv-LSTM is an extension of fully connected LSTM (FC-LSTM) by having convolutional structures for LSTM gating in both the input-to-state and state-to-state transitions (Xingjian et al., 2015). In Conv-LSTM, an extra connection with the previous memory cells is established to account for the effect of the previous input in the current timestamp (Xingjian et al., 2015).

In the training process, the memory cell can consider the effect of the earliest stages

(Rahman and Adjeroh, 2019). The main difference between CNN-LSTM and Conv-LSTM is that in CNN-LSTM, LSTM interprets the output from CNN model but in Conv-LSTM, the convolutions are used directly as part of reading input into LSTM (Brownlee, 2018). Conv-LSTM is suitable for predictions on 3-dimensional data such as spatiotemporal data.

The overall architecture of the model used in this study is presented in Figure 2 (c), and its detail is presented in Figure 5. In this study, a special form of Conv-LSTM, so called Conv-LSTM 2D provided by Keras library (a free open source library in Python), which combines gating of LSTM with 2D convolutions, was used. The functionality of Dropout and Dense layers is similar to what was described for the DNN and CNN-LSTM models.

5. Testing and demonstration

The proposed method was implemented on a live demolition project where a Komatsu PC220LC Hydraulic Excavator was in use. A commercial mobile app was used to collect noise level and IMU data including accelerometer, gyroscope and linear acceleration data in three-dimensional axes. Two android smartphones were mounted inside the cabin of the excavator on the window to mitigate the risk of losing data due to the risk of the app crashing on one phone or other incidents. The frequency of data capturing was 8 Hz, which was the highest frequency while the commercial app could run and capture data without crashing. Three hours of the excavator operation were monitored using the camera, IMU and noise sensors. Figure 6 shows a snapshot of the site and equipment involved.

Google Colab was used to train the models. Google Colab provides different computing resources, including Central processing units (CPU) and Graphics processing unit (GPU), which is faster than CPU. The computational time for the highest number of features (i.e. 10 features) was measured using both CPU and GPU. The results in Table 1 shows CPU-based training is slower than GPU-based training, but the maximum run time is less than 5 minutes.

As mentioned in the methodology section, this study intends to automatically measure the utilization ratio by recognizing two states of the equipment: active and inactive. During the monitored time, the excavator was mostly active working on demolishing a building. There were some occasions that the excavator operator stopped working for a short period of time, which was considered inactive time. Figures 7 to 10 show a sample of accelerometer, gyroscope, linear acceleration and noise level data when the excavator was active and inactive. The collected data were then pre-processed. The sliding window size for activity recognition was five seconds and was adequate because only two states (i.e active and inactive), not of complex nature, are involved. Since the frequency of data was 8 Hz, 40 data sets were available for each window. These data sets were labelled using the captured video.

Three deep learning models including DNN, CNN-LSTM and Conv-LSTM algorithms were created using Keras deep learning package with TensorFlow as a backend engine. The models were created for three combinations of data:

- Accelerometer and gyroscope data
- Accelerometer, gyroscope and linear acceleration data
- Accelerometer, gyroscope data, linear acceleration and noise level data

The train/test ratio of 75/25 was used for splitting the dataset into train and test sets in a stratified sampling fashion. Then, 80% of the train dataset was used as the actual train set and the remaining 20% was used as the validation set. Stratified sampling is based on splitting a data set in a way that each train and test subset has the same percentage of the samples from the complete target class. Therefore, this method can result in the training and test subsets with the input dataset that has the same ratio of the class labels. Thereafter, the model is iteratively trained and validated on these different sets.

The accuracy of the models was calculated based on the number of correct predictions and the total number of predictions as shown in Equation 4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

Table 1 shows the accuracy of experimented models for activity recognition. Figures 11, 12 and 13 display the validation loss and validation accuracy of the models containing the three combination inputs, (i.e., 6, 9 and 10 features), respectively.

The actual utilization ratio was measured manually as 89.85% using the captured video. Using the result of the models, the utilization ratio can be automatically calculated as shown in Table 3, and accuracy of the models was calculated by comparing it with the actual utilization ratio as shown in Table 4. As seen in Table 4, the accuracy of these combinations is very high (more than 97%) and the changes in absence of one input type (e.g. noise) are insignificant. Therefore, in the absence of some input data, the proposed approach is still applicable and provides a high accuracy.

6. Discussion

The developed model resulted in high accuracy for both activity recognition and productivity

doi: 10.1680/jsmic.21.00031

measurement. This accuracy can be attributed to capabilities of deep learning algorithms in feature engineering when a large amount of data is available. Another contributor to this high accuracy comparing to similar studies (e.g., Ahn et al. (2015) with 93% accuracy and Kim et al. (2018) with 91.83% accuracy), is the lower level of detail required for productivity measurement as this study considered two states (active and inactive) for the activity recognition. Although the accuracy levels are very high, they can be improved further with a larger amount of training data by increasing the frequency and/or duration of data collection. In this case study, DNN model using accelerometer and gyroscope data led to the highest accuracy (97.25%) for activity recognition. For productivity measurement, DNN using accelerometer, gyroscope data, linear acceleration and noise level data achieved the highest accuracy (99.78%). However, the variations of the achieved accuracies are insignificant among the models and the input combinations (less than 1% for activity recognition and less than 3% for productivity measurement), which could be explained by the low level of detail required for predictions. In addition, the results show that if some input data such as noise level is not available, the proposed approach can still provide a high accuracy with the other input features. If a higher level of details is considered, more variation could be observed to be able to compare capabilities of different algorithms and the impact of input data. In this case study, the excavator was doing only one type of activity (i.e., building demolition). However, for an excavator performing most or all of the steps involved in a round trip or cycle (i.e. pre-excavate, excavate, lift, unload, swing), the same approach proposed in this paper could be applied to recognise each step within the round trip provided that sufficient data is capture

doi: 10.1680/jsmic.21.00031

about each of the steps. In such cases, the combination inputs and the different algorithms are likely to perform differently. Therefore, to enhance the application of the model, other types of activities such as excavation and loading can be studied to make the model more generic for excavator operations.

A direct comparison between the proposed approach and other relevant methods available in the literature is shown in Table 5. It shows that the accuracy achieved by this method is higher than that of other studies using data from either smartphone sensors, accelerometers, or video surveillance. For instance, the study by Chen et al. (2020) used a vision-based method and could achieve 93.8% accuracy for measuring idle time, which is comparable with the accuracy achieved for the prediction of the inactive state in this study. Moreover, the main advantage of this method over other methods (e.g., vision-based and sensor-based methods) is in its low computational cost and scalability for large infrastructure sites due to its ability to cover a wide range of equipment and its portability, which are two main features that are not possible with fixed location sensors such as video surveillance.

7. Limitations and Future work

This study has some limitations that can be addressed in future. In the case study, the excavator activities were limited to demolition tasks. The proposed method can be experimented for other types of excavator activities such as excavation and loading to further substantiate its capabilities. As such, more complex activity recognition with more states may be required to study more detailed equipment operation efficiency.

The study measured the Utilisation Ratio which is a time-based productivity metric.

doi: 10.1680/jsmic.21.00031

Productivity can be also expressed as a production rate (e.g. measured as output of soil excavated). This has not been considered in this study as it would require data about the full operating cycle of an excavator (i.e. load, swing, dump, return). Other studies (i.e. Kassem et al., 2020) have measured this metric using telematics data but it revealed challenging to achieve high level of accuracy.

In this study, a commercial mobile application was used to capture the data. The application had limitations on the frequency for data capture. The highest frequency rate to avoid crashing was 8 Hz while in similar studies higher rates were used. Despite this limitation, the accuracy of the model in the case study was very high. In future studies, the impact of data capture frequency on the accuracy of the model can also be explored. In addition, performance and efficiency of this method can be further explored by experimenting other types of equipment such as loaders and cranes.

8. Conclusion

In this paper, a deep learning method was proposed for automating equipment productivity measurement. This method uses kinematic and noise level data captured by smartphone sensors. Three deep learning algorithms including DNN, CNN-LSTM, and Conv-LSTM were investigated for activity recognition of an excavator and productivity measurement.

The results of the experiment showed high accuracy of the models (over 96.70% for activity recognition and over 97.22% for productivity measurement). Equipment-intensive construction projects can benefit from the proposed method to automate productivity measurement, identify equipment inefficiencies in near real-time, and inform corrective actions

Accepted manuscript
doi: 10.1680/jsmic.21.00031

in lagging-behind performance of certain site zones. The findings can also inform the establishment of performance benchmarks for earthwork equipment. Such benchmarks can be built over time from several projects and can inform project budgeting and allocation of equipment, hence contributing to the resolution of the equipment redundancy problem faced in many large construction projects (Kassem et al., 2021).

References

Ahn, C. R., Lee, S. & Peña-Mora, F. Monitoring system for operational efficiency and environmental performance of construction operations using vibration signal analysis. Construction Research Congress 2012: Construction Challenges in a Flat World, 2012. 1879-1888.

Ahn, C. R., Lee, S. & Peña-Mora, F. 2015. Application of low-cost accelerometers for measuring the operational efficiency of a construction equipment fleet. Journal of Computing in Civil Engineering, 29, 04014042.

Akhavian, R. & Behzadan, A. H. 2015. Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers. Advanced Engineering Informatics, 29, 867-877.

Bae, J., Kim, K. & Hong, D. 2019. Automatic identification of excavator activities using joystick signals. International Journal of Precision Engineering and Manufacturing, 20, 2101-2107.

Banos, O., Galvez, J.-M., Damas, M., Pomares, H. & Rojas, I. 2014. Window size impact in human activity recognition. Sensors, 14, 6474-6499.

Brownlee, J. 2016. Crash Course in Recurrent Neural Networks for Deep Learning [Online]. Available: https://machinelearningmastery.com/crash-course-recurrent-neural-networks-deep-learning/ [Accessed December 18, 2020 2020].

Brownlee, J. 2018. LSTMs for Human Activity Recognition Time Series Classification [Online]. Available: https://machinelearningmastery.com/how-to-develop-rnn-models-for

-human-activity-recognition-time-series-classification/ [Accessed December 18, 2020 2020].

Bügler, M., Borrmann, A., Ogunmakin, G., Vela, P. A. & Teizer, J. 2017. Fusion of photogrammetry and video analysis for productivity assessment of earthwork processes. Computer-Aided Civil and Infrastructure Engineering, 32, 107-123.

Chen, C., Zhu, Z. & Hammad, A. 2020. Automated excavators activity recognition and productivity analysis from construction site surveillance videos. Automation in Construction, 110, 103045.

Cheng, C.-F., Rashidi, A., Davenport, M. A. & Anderson, D. V. 2017. Activity analysis of construction equipment using audio signals and support vector machines. Automation in Construction, 81, 240-253.

Cheng, C.-F., Rashidi, A., Davenport, M. A. & Anderson, D. V. 2019. Evaluation of software and hardware settings for audio-based analysis of construction operations. International Journal of Civil Engineering, 17, 1469-1480.

Chollet, F. 2017. Deep learning with Python. New York: Manning Publications Company.

Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K. & Darrell, T. Long-term recurrent convolutional networks for visual recognition and description. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. 2625-2634.

Gerami Seresht, N. & Fayek, A. R. 2018. Dynamic modeling of multifactor construction productivity for equipment-intensive activities. Journal of Construction Engineering and

Management, 144, 04018091.

Gong, J. & Caldas, C. H. 2011. An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations. Automation in Construction, 20, 1211-1226.

Gurmu, A. T. & Aibinu, A. A. 2017. Construction equipment management practices for improving labor productivity in multistory building construction projects. Journal of Construction Engineering and Management, 143, 04017081.

Huang, C.-J. & Kuo, P.-H. 2018. A deep cnn-lstm model for particulate matter (PM2. 5) forecasting in smart cities. Sensors, 18, 2220.

Ibbs Jr, C. W. & Terveer, K. R. 1984. Integrated construction preventive maintenance system. Journal of construction engineering and management, 110, 43-59.

Ibrahim, M. & Moselhi, O. 2014. Automated productivity assessment of earthmoving operations. Journal of Information Technology in Construction (ITcon), 19, 169-184.

Jagushte, R.V. 2017. Usability review of telematics for construction equipment fleet management. Master Thesis. University of Florida. Available at: https://ufdc.ufl.edu/UFE0051120/00001

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. & Huntingdon, J. 2021. Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach. Automation in Construction, 124, 103532.

Kassem, M., Rogage, K., Huntingdon, J., Durojaye, G., Arena, N., Kelly, G., Lund, T, and Clarke, T. 2018. Measuring and improving the productivity of construction's site equipment Fleet: An integrated IoT and BIM system. In Proceedings of the 36th International Conference of CIB W78, Newcastle-Upon-Tyne, UK, 18–20 September (2019), pp. 901-911, ISSN: 2706-6568.

Kim, H., Ahn, C. R., Engelhaupt, D. & Lee, S. 2018. Application of dynamic time warping to the recognition of mixed equipment activities in cycle time measurement. Automation in Construction, 87, 225-234.

Kim, J. & Chi, S. 2020. Multi-camera vision-based productivity monitoring of earthmoving operations. Automation in Construction, 112, 103121.

Mahmud, T., Sayyed, A. S., Fattah, S. A. & Kung, S.-Y. 2020. A Novel Multi-Stage Training Approach for Human Activity Recognition from Multimodal Wearable Sensor Data Using Deep Neural Network. IEEE Sensors Journal, 21, 1715-1726.

Montaser, A. & Moselhi, O. RFID+ for tracking earthmoving operations. Construction Research Congress 2012: Construction Challenges in a Flat World, 2012. 1011-1020.

Montaser, A. & Moselhi, O. 2014. Truck+ for earthmoving operations. Journal of Information Technology in Construction (ITcon), 19, 412-433.

Mutegeki, R. & Han, D. S. A CNN-LSTM approach to human activity recognition. 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020. IEEE, 362-366.

Nepal, M. P. & Park, M. 2004. Downtime model development for construction equipment management. Engineering, Construction and Architectural Management.

Ok, S. C. & Sinha, S. K. 2006. Construction equipment productivity estimation using artificial neural network model. Construction Management and Economics, 24, 1029-1044.

Rahman, S. A. & Adjeroh, D. A. 2019. Deep learning using convolutional LSTM estimates biological age from physical activity. Scientific reports, 9, 1-15.

Rashid, K. M. & Louis, J. 2019. Times-series data augmentation and deep learning for construction equipment activity recognition. Advanced Engineering Informatics, 42, 100944.

Sabillon, C., Rashidi, A., Samanta, B., Davenport, M. A. & Anderson, D. V. 2020. Audio-based bayesian model for productivity estimation of cyclic construction activities. Journal of computing in civil engineering, 34, 04019048.

Swapna, G., Kp, S. & Vinayakumar, R. 2018. Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals. Procedia computer science, 132, 1253-1262.

Vorster, M. C. & De La Garza, J. M. 1990. Consequential equipment costs associated with lack of availability and downtime. Journal of Construction Engineering and Management, 116, 656-669.

Wang, J., Chen, Y., Hao, S., Peng, X. & Hu, L. 2019. Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters, 119, 3-11.

Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K. & Woo, W.-C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems, 2015. 802-810.

Table 1. Comparative analysis between CPU and GPU performance

Computing resource in Colab	DNN	CNN-LSTM	CONV-LSTM
CPU	1’ 20’’	4’ 10’’	3’ 55’’
GPU	58’’	2’ 56’’	2’ 8’’

doi: 10.1680/jsmic.21.00031

Table 2. Accuracy for activity recognition

Input Data	Model Accuracy (%)		
	DNN	CNN-LSTM	Conv-LSTM
Accelerometer and gyroscope data	97.25%	96.93%	96.85%
Accelerometer, gyroscope and linear acceleration data	97.17%	96.77%	97.01%
Accelerometer, gyroscope data, linear acceleration and noise level data	97.01%	96.70%	97.01%

doi: 10.1680/jsmic.21.00031

Table 3. Predicted utilization ratio

Input Data	Predicted Utilization Ratio (%)		
	DNN	CNN-LSTM	Conv-LSTM
Accelerometer and gyroscope data	91.08%	91.36%	90.13%
Accelerometer, gyroscope and linear acceleration data	91.26%	92.34%	90.17%
Accelerometer, gyroscope data, linear acceleration and noise level data	90.04%	91.37%	91.45%

doi: 10.1680/jsmic.21.00031

Table 4. Accuracy for measuring utilization ratio

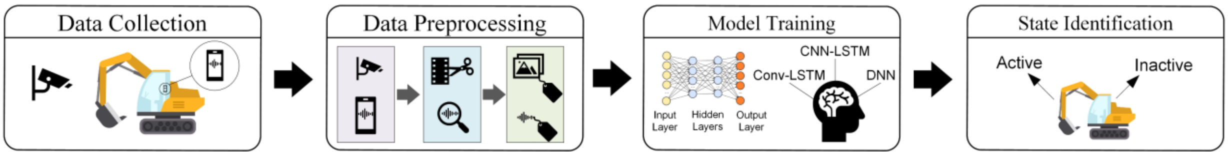
Input Data	Utilization Ratio Accuracy (%)		
	DNN	CNN-LSTM	Conv-LSTM
Accelerometer and gyroscope data	98.63%	98.31%	99.68%
Accelerometer, gyroscope and linear acceleration data	98.43%	97.22%	98.53%
Accelerometer, gyroscope data, linear acceleration and noise level data	99.78%	99.42%	98.21%

Accepted manuscript
doi: 10.1680/jsmic.21.00031

Table 5. comparison between the results of this study and some other studies.

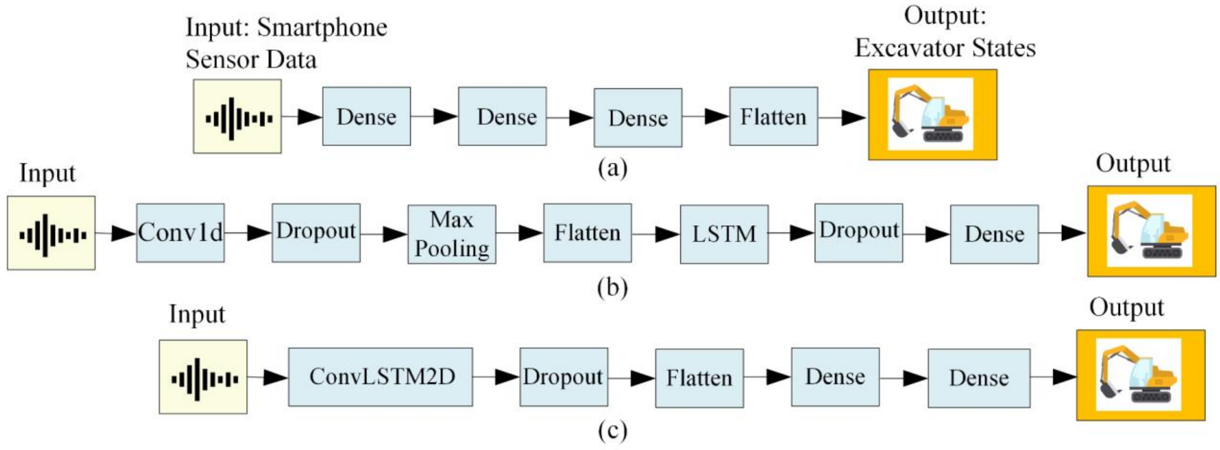
Reference	Used Data Type	Used Algorithms	Main Parameters	Accuracy	Number of states
Ahn et al. (2015)	Accelerometer data	Naïve Bayes, Instance-based learning, K-nearest neighbor (KNN) and Decision tree (J48)	Frequency: 100 Hz Sliding window size: 128-sample windows	93% for the activity classification	3
Kim et al. (2018)	Smartphone sensors (IMU)	Random Forest, Naïve Bayes, J48, and SMO	Frequency: 128 Hz Sliding window size: 1 second	91.83% accuracy for cycle time measurement	3
Cheng et al. (2020)	Surveillance video data	Faster R-CNN for excavator detection and deep Simple Online and Real-Time (SORT) for excavator tracking	Frequency: 25 frames per second (FPS), Sliding window size: 4 seconds.	93.8% for idle time measurement	2
This study	Smartphone sensor data	DNN, CNN-LSTM and Conv-LSTM were compared	Frequency: 8 Hz Sliding window size: 5 seconds.	99.78%for utilization ratio measurement	2

Figure 1. Modelling process



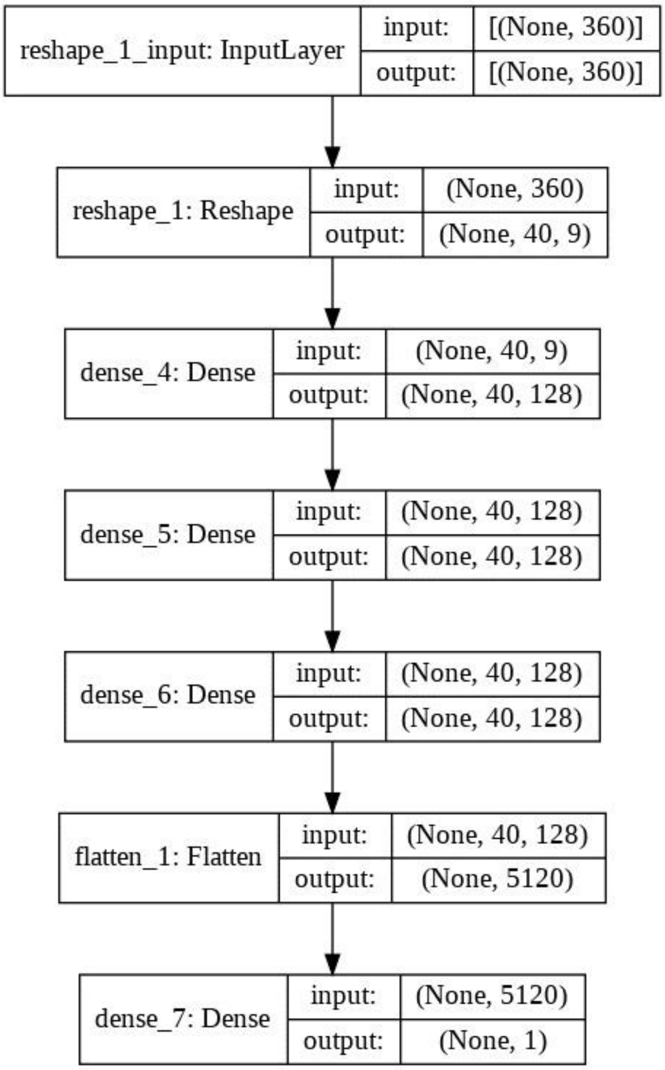
Accepted manuscript
doi: 10.1680/jsmic.21.00031

Figure 2. (a) DNN model architecture; (b) CNN-LSTM model architecture; (c) Conv-LSTM model architecture



doi: 10.1680/jsmic.21.00031

Figure 3. Detail architecture of DNN model for 9 features



Accepted manuscript
doi: 10.1680/jsmic.21.00031

Figure 4. Detail architecture of CNN_LSTM model for 9 features

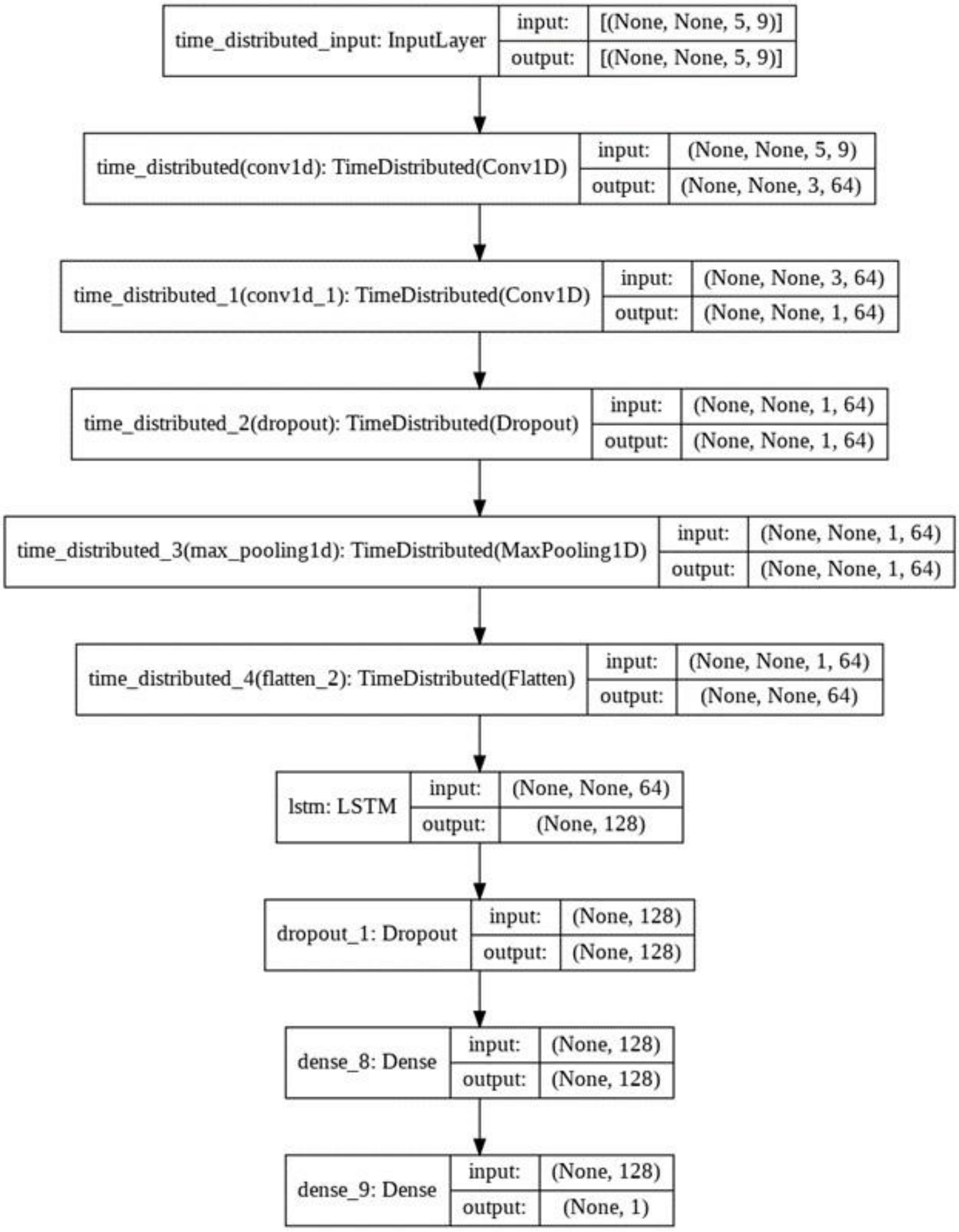
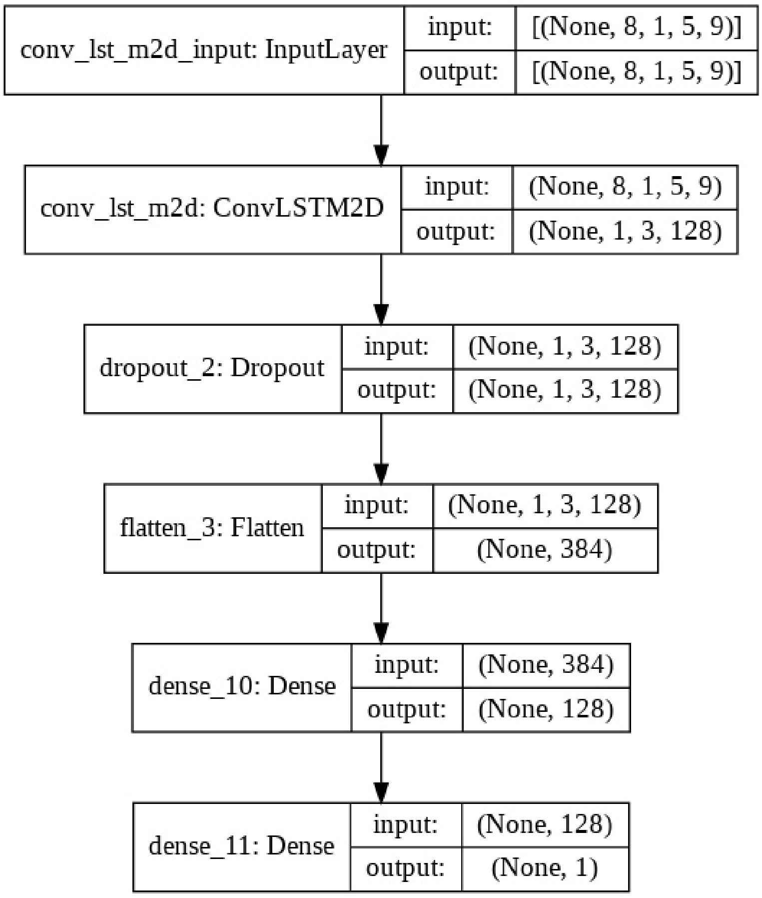


Figure 5. Detail architecture of CONV_LSTM model for 9 features

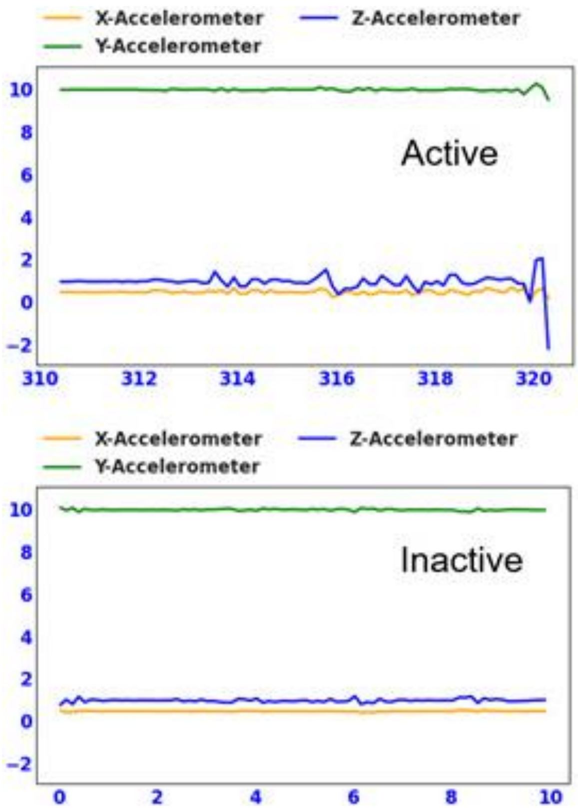


Accepted manuscript
doi: 10.1680/jsmic.21.00031

Figure 6. A snapshot of the captured video

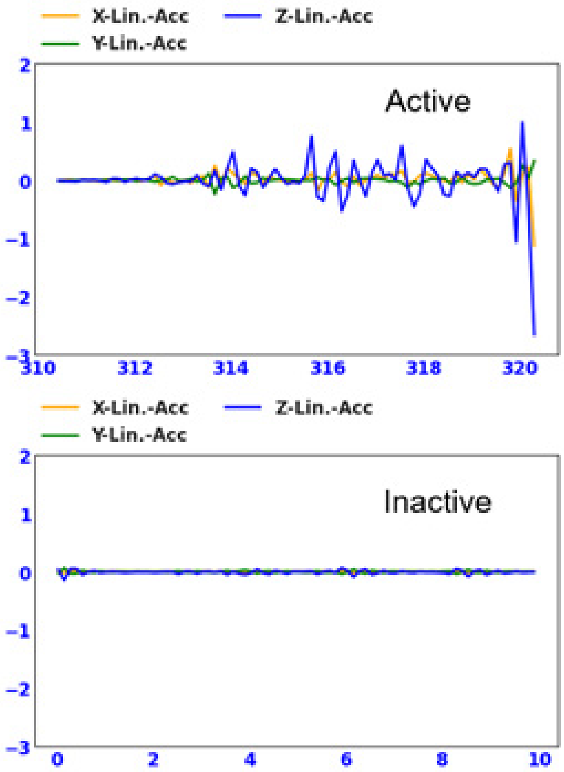
Accepted manuscript
doi: 10.1680/jsmic.21.00031

Figure 7. Sample of accelerometer data in x, y and z axis for active and inactive states



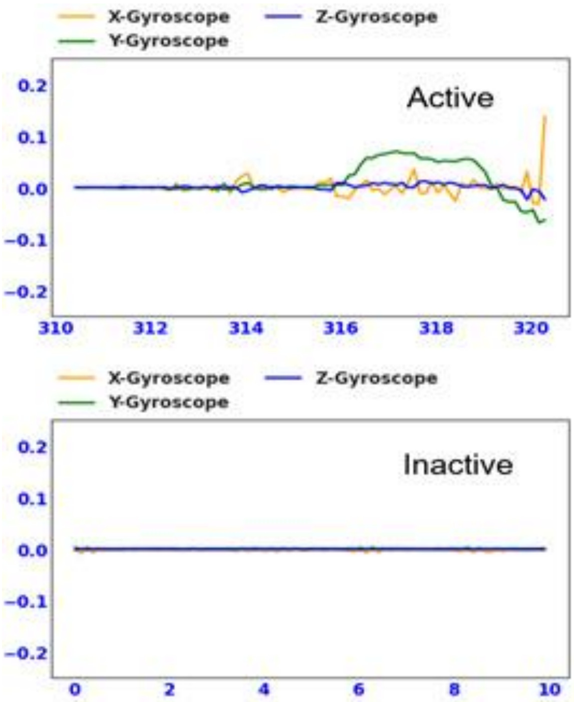
Accepted manuscript
doi: 10.1680/jsmic.21.00031

Figure 8. Sample of linear accelerometer data in x, y and z axis for active and inactive states



Accepted manuscript
doi: 10.1680/jsmic.21.00031

Figure 9. Sample of gyroscope data in x, y and z axis for active and inactive states



Accepted manuscript
doi: 10.1680/jsmic.21.00031

Figure 10. Sample of noise level data for active and inactive states

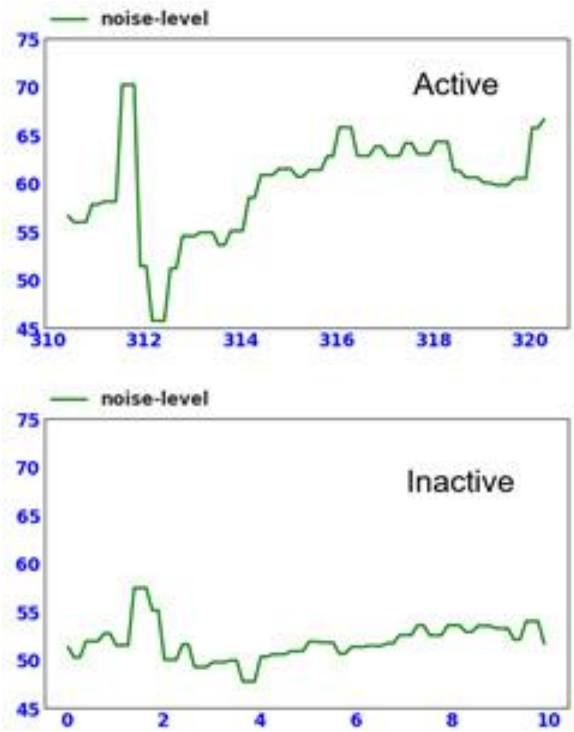
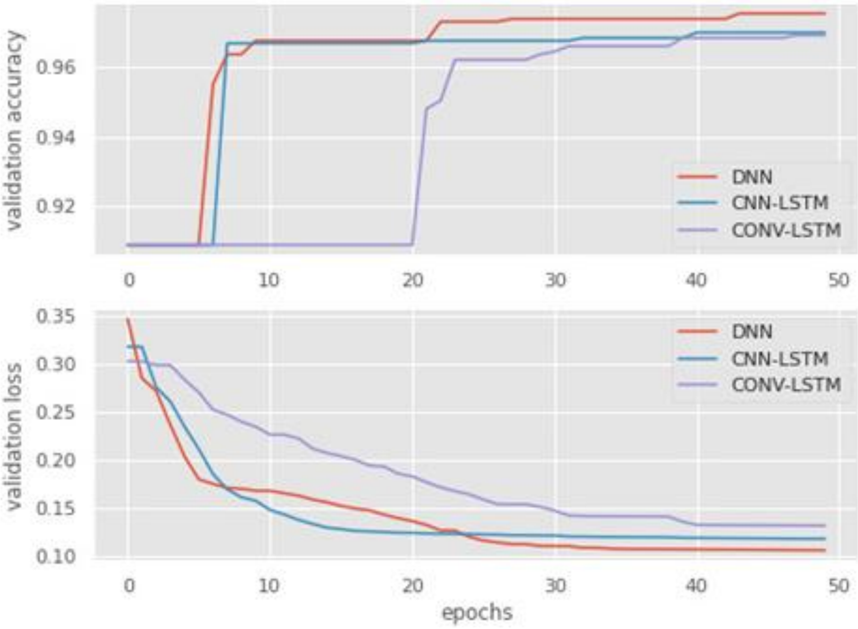


Figure 11. validation loss and validation accuracy of models with 6 features as input data (Accelerometer and gyroscope data)



Accepted manuscript
doi: 10.1680/jsmic.21.00031

Figure 12. validation loss and validation accuracy of models with 9 features as input data (Accelerometer, gyroscope and linear acceleration data)

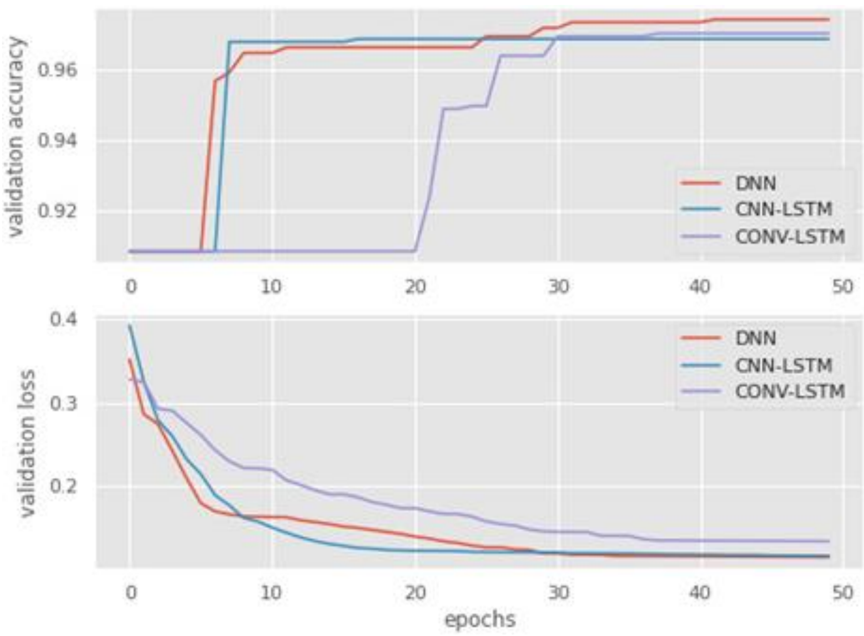


Figure 13. validation loss and validation accuracy of models with 9 features as input data (Accelerometer, gyroscope data, linear acceleration and noise level data)