www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

RELEVANCE OF DEEP SEQUENCE MODELS FOR RECOGNISING AUTOMATED CONSTRUCTION ACTIVITIES: A CASE STUDY ON A LOW-RISE CONSTRUCTION SYSTEM

SUBMITTED: March 2023
REVISED: August 2023
PUBLISHED: August 2023
EDITOR: Robert Amor
DOI: 10.36680/j.itcon.2023.023

Aparna Harichandran, Postdoctoral fellow
Department of Civil and Architectural Engineering, Aarhus University, Denmark
E-mail: aparnahari@alumni.iitm.ac.in

Benny Raphael, Professor
Department of Civil Engineering, Indian Institute of Technology Madras, India
E-mail: benny@civil.iitm.ac.in

Abhijit Mukherjee, Professor
School of Civil and Mechanical Engineering, Curtin University, Australia
E-mail: abhijit.mukherjee@curtin.edu.au

SUMMARY: *Recognising activities of construction equipment is essential for monitoring productivity, construction progress, safety, and environmental impacts. While there have been many studies on activity recognition of earth excavation and moving equipment, activity identification of Automated Construction Systems (ACS) has been rarely attempted. Especially for low-rise ACS that offers energy-efficient, cost-effective solutions for urgent housing needs, and provides more affordable living options for a broader population. Deep learning methods have gained a lot of attention because of their ability to perform classification without manually extracting relevant features. This study evaluates the feasibility of deep sequence models for developing an activity recognition framework for low-rise automated construction equipment. Time series acceleration data was collected from the structure to identify major operation classes of an ACS. Long Short Term Memory Networks (LSTM) were applied for identifying the activity classes and the performance was compared with that of traditional machine learning classifiers. Diverse augmentation methods were adopted for generating datasets for training the deep learning classifiers. Several recently published literature seem to establish the superiority of complex deep learning techniques over traditional machine learning algorithms regardless of the application context. However, the results of this study show that all the conventional machine learning classifiers perform equivalently or better than deep learning classifiers in identifying activities of the ACS. The performance of the deep learning classifiers is affected by the lack of diversity in the initial dataset. If the augmented dataset significantly alters the characteristics of the original dataset, it may not deliver good recognition results.*

KEYWORDS: *Automated activity recognition, Deep learning, Machine learning, Construction monitoring, Data augmentation.*

REFERENCE: *Aparna Harichandran, Benny Raphael, Abhijit Mukherjee (2023). Relevance of deep sequence models for recognising automated construction activities: a case study on a low-rise construction system. Journal of Information Technology in Construction (ITcon), Vol. 28, pg. 458-481, DOI: 10.36680/j.itcon.2023.023*

1. INTRODUCTION

Automated construction has several promising characteristics compared to conventional construction such as higher speed of building, lower lifecycle costs, better building quality, and fewer labour costs (Castro-Lacouture, 2009). Nevertheless, the adoption of automated construction techniques is still in the early stages. Although, most of the existing Automated Construction Systems (ACS) are for high-rise buildings (Bock and Linner, 2016), changing socioeconomic circumstances and emerging industrialised construction technologies gravitate towards the adoption of low-rise ACS (Harichandran *et al.*, 2023). According to the latest accident statistics, occupational accidents and deaths are most common in the construction sector (Bureau of Labor Statistics, 2021). The complex nature of a construction site that comprises various tasks, resources and workers contributes to the high incident rate. Some of the safety risks in conventional construction can be reduced by deploying ACS. However, human-machine interaction in the context of automated construction unfolds yet another complex scenario and necessitates the implementation of a robust monitoring system. Recognising construction activities is an essential first step for developing such a monitoring system.

The activities of construction equipment are identified for several objectives including calculating cycle times for operations, estimating fuel consumption, emission rates, or equipment productivity, evaluating the state of the equipment, and monitoring the progress of the construction projects (Harichandran *et al.*, 2020a; Rashid and Louis, 2020; Sherafat *et al.*, 2022). Traditional machine learning methods require manually extracting features from raw data to train classification models. For automatically classifying equipment actions without manually extracting features, advanced deep learning techniques are frequently adopted (Chen *et al.*, 2023). By directly learning high-level features from unprocessed sequential input, recurrent neural networks (RNNs) are able to perform intricate activity recognition (Rashid and Louis, 2019). Large data sets are frequently needed for deep sequence model training; for this reason, synthetic data is produced using a variety of data augmentation techniques (Forestier *et al.*, 2017). Kinematic measurements such as location or time series vibration data from the equipment were extracted for developing an efficient activity recognition framework (Kim *et al.*, 2021; Langroodi *et al.*, 2021).

Several studies on construction equipment activity recognition advocate deep learning models based on initial evaluation results (Scarpiniti *et al.*, 2021; Sherafat *et al.*, 2022; Chen *et al.*, 2023). However, the implementation of deep learning techniques often requires high computational resources and time. Besides, the initially defined activity recognition problem experiences data drift and concept drift over the life cycle of a project (Hu *et al.*, 2020). Collection of additional data and retraining the models are necessary for that context and are often more resource-intensive for deep learning models. Therefore, the selection of advanced models should be justified by their performance proportional to the resources needed for a specific project.

Even though the existing studies on construction equipment activity recognition show promising results, they are mainly focused on earth excavation and moving equipment. Methods for identifying the activities of ACS are limited and the identification problem involves unique challenges. The complexity of the problem arises from the subtlety of movements and activities in ACS, which often do not have clear articulating parts like excavators. Identifying distinct activities in such a system can be quite challenging. The ACS performs coordinated movements to accomplish specific tasks, and even small variations in these movements can denote different activities. Therefore, pattern recognition problems for ACS activities are much more complex compared to other construction equipment with distinct activity patterns. Consequently, an activity recognition framework that specifically addresses these challenges needs to be developed and deep sequence models show potential for this problem.

This study evaluates the feasibility of deep sequence models for developing an activity recognition framework for automated construction equipment through a case study on a low-rise ACS. To determine the main operating classes of an ACS, time series acceleration data from the structure during the operation was gathered. In order to categorize the activities, Long Short Term Memory Networks (LSTM) were deployed. Diverse augmentation methods were adopted for generating datasets for training the deep learning classifiers and their performance was compared to that of conventional machine classifiers. The acceleration data during automated construction was collected from a real-world scale ACS. The results of this study contribute to developing a robust activity recognition framework and monitoring system for ACS. Besides, the study provides valuable insights into the influence of data augmentation methods on activity recognition performance.

The remaining sections of the paper are organized as follows. Section 2 provides the background of ACS and a review of studies on construction equipment activity recognition. Section 3 presents the research methodology and section 4 describes the automated top-down construction method and the case study. The results and discussion are presented in section 5. The conclusions drawn from this study are given in section 6. Finally, the limitations of this study and the outlook are presented in section 7.

2. BACKGROUND AND RELATED WORK

2.1. Automated Construction Systems

Even though automated construction systems are not widely adopted in the construction industry today, it has been successfully implemented in several high-rise construction projects in Japan (Cai *et al.*, 2019). Bock and Linner present a comprehensive analysis of ACS for high-rise buildings and classify them based on the construction scheme (Bock and Linner, 2016). The main operation unit of the ACS is referred to in the literature as a 'factory'. If the factory is located at ground level during construction, it is called a ground factory. In this system, each floor of the building is assembled at ground level and lifted up while the factory remains fixed in the ground (Sekiguchi *et al.*, 1997). If the factory is placed on the top of the building under construction and sequentially lifted with the progress of construction, it is termed a sky factory (Yamazaki and Maeda, 1998; Wakisaka *et al.*, 2000).

The ground factory systems are categorised into three: 1) fixed ground factory that pushes the buildings up, 2) combined on-site and off-site factory, and 3) horizontally moving self-supported factory (Bock and Linner, 2016). The construction progress from the top floor to the bottom floor in ground factory systems that have the building push-up method. The floors will be completed at the ground level and lifted upwards while the ground factory remains in a fixed location. The orientation of the buildings is vertical in high-rise buildings (for example, Automatic Up-Rising Construction by Advanced Technique (AMURAD) (Sekiguchi *et al.*, 1997)) and horizontal in low-rise buildings (for example, System Skanska and J-up (Bock and Linner, 2016)). The second category of ground factory systems implemented for low-rise construction is combined on-site and off-site factory systems. NCC Komplett, developed by NCC, Sweden belongs to this category; this system involves the synchronised operation of two factories (Bock and Linner, 2016). The on-site factory consists of a self-supporting hall structure that provides weather protected environment and subsystems for handling and connecting high-level building components. The off-site factory prefabricates and finishes concrete building components and transports them to the on-site factory through delivery trucks. The third ground factory system, the horizontally moving self-supported factory, is mainly developed for long horizontally oriented buildings. The ground factory covers the structure and moves horizontally on a rail with construction progress. These systems belong to the mechanized category rather than automated systems. Some examples of these systems include Bauhelling Summerfield developed by AHAG-Sommerfeld, Germany and; Bauschiff developed by Neufert, Germany (Bock and Linner, 2016).

The majority of the ACS implemented in high-rise construction belong to the sky factory systems. These systems follow a variety of construction schemes. In most sky factory systems, the factory is supported by the building under construction and moves upwards as the work progresses. Automated structural steel Building Construction System (ABCS) developed by Obayashi, Japan (Wakisaka *et al.*, 2000) and Shimizu Manufacturing System by Advanced Robot Technology (SMART) developed by Shimizu, Japan (Yamazaki and Maeda, 1998) are some of the examples. In another construction scheme, the sky factory is supported by stilts of its own, independent of the building structure. The sky factory provides a weatherproof working environment like the earlier construction scheme. However, the synchronisation of construction work was simplified since the sky factory moves upwards on the extending stilts instead of being supported by the structure. BIG CANOPY developed by Obayashi, Japan (Hamada *et al.*, 1998) is an example. The third construction scheme involves a sky factory and a core factory, both moving upwards with the construction progress. The sky factory is pulled upwards along the core structure, which is built in advance by the core factory. Robotic and Crane based Automatic Construction System (RCACS) developed by the Korean Consortium, South Korea, belong to this construction scheme (Kang *et al.*, 2011). The core factory with limited functionality follows a simple construction scheme for building the structural core. The main sky factory deals with significant construction operations. Other categories of sky factory systems include a combination of conventional construction and centralized or decentralized sky factories (Bock and Linner, 2016).

The current Automated Construction Systems (ACS) around the world are primarily intended for the construction of high-rise structures (Wakisaka *et al.*, 2000; Gassel, 2005; Bock and Linner, 2016). The scarcity of automated systems specifically designed for low-rise structures exposes a significant gap in the construction industry. The introduction of low-rise ACS has the potential to significantly accelerate construction processes, reduce costs, improve quality, and contribute to sustainability. Automating the construction of low-rise buildings is being acknowledged in the changing socioeconomic circumstances. The best possible housing solutions for people affected by natural calamities or a large number of temporary testing facilities during pandemics (e.g. COVID-19) can be quickly provided through construction automation (Harichandran *et al.*, 2023). Furthermore, the implementation of low-rise ACS encourages energy-efficient construction practices, optimized material usage, and waste reduction, aligning with global sustainability goals. By streamlining construction processes and leveraging automation for precision and efficiency, low-rise ACS offers a cost-effective solution that can translate into more affordable housing options, thereby supporting more equitable community development. Nevertheless, the technology for developing low-rise ACS is sparingly explored. The comprehensive review of ACS presented by Bock and Linner includes thirty systems, merely five of which are designed for low-rise structures (Bock and Linner, 2016). Out of these five systems, two are historical mechanized construction prototypes rather than automated systems. The other three low-rise ACS are NCC Komplett, J-up and System Skanska. Even though these systems possess automated subsystems for various operations, they lack a real-time monitoring system. An automated top-down construction system for low-rise structures has been developed by the authors of this study along with others through various laboratory prototypes (Raphael *et al.*, 2016; Harichandran *et al.*, 2020b, 2021, 2023). The purpose of this study is to contribute towards developing an integrated automated monitoring system for the ACS.

2.2. Equipment activity recognition in construction

Activity recognition of humans or equipment is an area of interest for many disciplines including health science, ergonomics and construction management. Equipment activity recognition in construction is interesting because it helps to formulate tangible and reliable indices to evaluate the overall performance of a construction project. Many studies have focused on identifying activities of equipment deployed on earth excavation and moving. The existing studies in equipment activity recognition can be broadly categorized based on the data acquired, which include images/videos, sound, and kinematic measurements.

2.2.1. Computer vision-based methods

Computer vision-based equipment activity identification studies were aided by easy access to powerful computing facilities and affordable storage devices. From video footages of the construction site, Golparvar-Fard et al. classified the activities of the excavator and truck using Support Vector Machines (SVM) by representing the spatio-temporal visual features into Histogram of Oriented Gradients (HOG) (Golparvar-Fard *et al.*, 2013). Kim et al. analysed the interaction of earth excavation and moving equipment during their operation cycle and introduced that into the activity recognition problem along with a proximity threshold (J. Kim *et al.*, 2018). Images were used to identify the construction activities through the Tracking-Learning-Detection (TLD) method. A similar study conducted by Kim and Chi used sequential patterns of visual features and operation cycles to recognise the activities of an excavator (Kim and Chi, 2019). They have used TLD and a hybrid of Convolutional Neural Network (CNN) and Double-layer Long Short Term Memory (LSTM) networks only to result in an average accuracy of 93.8%. Chen et al. detect activities of multiple excavators from long site surveillance videos (Chen *et al.*, 2020). Spatio-temporal features were extracted from the videos and faster region proposal convolutional neural network (R-CNN) was used to identify three operation classes with an overall accuracy of 87.6%. In a recent study, Chen et al. adopted zero-shot learning for identifying the activities of the excavator and loader (Chen *et al.*, 2023). In this study, equipment detection, tracking and activity recognition are performed by You Look Only Once (YoloV5), Simple Online and Real-Time Tracking (SORT) and Contrastive Language Image Pre-training (CLIP) respectively. Ghelmani and Hammad proposed CVRLoLD (Contrastive Video Representation Learning on Limited Dataset), a self-supervised contrastive learning method for construction equipment activity recognition with limited labelled data. (Ghelmani and Hammad, 2023). By training a backbone network on unlabelled data and fine-tuning it with labelled data, the proposed method achieves an 81.7% accuracy in recognizing activities using only 30% of the dataset's labels. This approach shows the possibility to reduce data labelling time and effort while maintaining good performance with the construction industry's limited datasets. The limitations of the visual data-based methods include the need for a favourable surrounding environment and constraints for identifying activities of equipment with little movement or no articulating parts.

2.2.2. Audio-based methods

Audio-based methods are particularly suitable for identifying equipment that emits distinctive sounds while operating (Sherafat *et al.*, 2020). Therefore, most of the research in the field focuses on equipment like excavators, hydraulic hammers, electric hammers, and cutting machines (Cao, Huang, *et al.*, 2017; Cao, Wang, *et al.*, 2017). The microphones can be installed inside the equipment cabin or close to the equipment to gather audio data. The capability of these techniques to identify various types of equipment is one of their main benefits. An example is the study by Cheng et al. that correctly identified eleven distinct machine types with an accuracy of over 80% (Cheng *et al.*, 2017). Rashid and Louis developed an automated activity identification framework for modular construction factories using sound as a data source (Rashid and Louis, 2020). They investigated the effects of various features extracted from different domains of audio signals (time, time-frequency, cepstral, and wavelet) on the performance of the activity identification model. By optimizing the feature space through sensitivity analyses and feature ranking techniques, a 130-dimensional feature vector with a 0.5-second window size is designed, achieving a high 97% F-1 score for identifying different activities. In a recent study, Sherafat et al. proposed a multi-label multi-level sound classification method for recognizing activities of heavy construction equipment using sound data. The method utilizes Short-Time Fourier Transform (STFT) and Convolutional Neural Network (CNN) with a single-channel microphone and includes data augmentation to simulate real-world equipment sound mixtures. The results indicate that the proposed method effectively identifies activities of multiple equipment pieces on construction job sites without the need for pre-separating sound signals, suggesting its potential for practical application in the construction industry (Sherafat *et al.*, 2022). However, the level of detail of activities identified by sound based methods is limited even with advanced features.

2.2.3. Kinematic data-based methods

Kinematic data-based methods for activity recognition mainly use data corresponding to the location, acceleration and spatial orientation of the equipment. Previous studies demonstrate that kinematic data-based approaches give excellent accuracy and a high degree of detail, even with cost-effective data collection methods involving low-cost accelerometers (Ahn *et al.*, 2015) and sensors built-in mobile phones (Akhavian and Behzadan, 2015). However, the performance of identification decreases when high-level details of activities are recognised. Kim et al. showed that the Dynamic Time Warping (DTW) method improves the identification performance of existing machine learning techniques (H. Kim *et al.*, 2018). Rashid and Louis implemented data augmentation methods for improving the performance matrices for identifying activities of a front-end loader and an excavator by training shallow networks (Artificial Neural Networks (ANN)) and deep neural networks (LSTM) using large datasets (Rashid and Louis, 2019). Shi et. al. show that through main pump pressure and displacement data, the working cycle stages of an excavator can be identified with an accuracy of 93.82% (Shi *et al.*, 2020). Simple machine learning algorithms were applied and the domain knowledge is introduced in the identification problem through an intelligent calibration system. Similar work that used acceleration data and an advanced deep learning method (a hybrid network of CNN and LSTM) resulted in an accuracy of over 77% for the excavator (Slaton *et al.*, 2020). Langroodi et al. proposed an approach that combines the Random Forest classifier with a fractional calculus-based feature augmentation technique, to create an accurate activity recognition model with limited data (Langroodi *et al.*, 2021). The findings demonstrate that the fractional feature augmentation technique improves the performance of various machine learning methods, including Neural Networks and Support Vector Machines and achieves comparable results to deep learning methods but with a significantly smaller training dataset. Meng and Zhu addressed the lack of efficient usage of vibration monitoring data in establishing an empirical vibration model for construction activities by proposing an activity recognition model that combines a convolutional neural network (CNN) and the RandAugment algorithm (Meng and Zhu, 2022). The results demonstrate that the well-trained CNN with RandAugment achieves a high accuracy of 99.21% in classifying construction activities, outperforming the multilayer perceptron (MLP) model. Harichandran et al. formulated the activity recognition problem considering the hierarchical relationship between activity classes and maintained the performance of identification at various levels of details of activities (Harichandran *et al.*, 2021). Subsequently, a hybrid unsupervised and supervised machine learning (HUS-ML) framework was introduced to recognise the activities and faulty conditions during automated construction (Harichandran *et al.*, 2023). Considering the major attributes of kinematic data-based methods such as high accuracy, the potential to identify a high level of details and the ability to capture equipment with limited movements, it is selected for identifying the activities of ACS. Since the ACS operations predominantly include characteristic vibrations in the structure, acceleration data is selected for activity identification.

2.3. Gaps in knowledge and point of departure

According to the literature review performed on Automated Construction Systems and activity identification of construction equipment, the following gaps in the knowledge have been identified.

- The current studies on construction equipment activity recognition are focused mainly on earth excavation and moving equipment where each activity has distinct patterns. Methods for identifying automated construction activities with subtle variations in pattern are limited.
- Most current studies involve manually extracting features from sensor data for activity classification.
- The effect of data augmentation techniques on the performance of classifiers in pattern recognition problems with limited datasets is sparsely explored.

Considering gaps drawn from the literature, this study aims to evaluate the feasibility of developing an activity recognition framework for low-rise ACS based on LSTM networks that can learn high-level features directly from the raw data. Diverse augmentation methods were adopted for generating datasets for training the deep learning classifiers and the influence of augmentation methods was evaluated. The performance of the deep sequence models was compared with conventional machine learning classifiers such as k-Nearest Neighbour (kNN), Decision Tree (DT), Support Vector Machines (SVM), Discriminant Analysis (DA), Naïve Bayes (NB), and Artificial Neural Network (ANN).

3. METHODOLOGY

The methodology adopted in this study for identifying automated construction activities is illustrated in Fig. 1, which consists of four principal stages: automated construction and data collection, data pre-processing and time series data augmentation, the training of deep sequence models and hyperparameter tuning, and finally, the evaluation of models and performance benchmarking. The first stage involves conducting construction operations through the Automated Construction System (ACS). The ACS's actions induce vibration on the structure being assembled, and this vibration is captured in the form of raw acceleration data using strategically placed accelerometers. This data serves as the baseline for our activity recognition problem, representing the unique signatures of different construction activities.

In the second stage, we pre-process this raw data to render it suitable for input into our learning models. Pre-processing involves eliminating noise and normalizing the data to achieve uniformity. Further, we perform time series data augmentation to expand our dataset, hence providing a more robust foundation for deep learning classification. Various data augmentation techniques, such as jittering, scaling, downsampling, and oversampling, are employed, each with its unique impact on the resultant learning models. A detailed description of time series data augmentation methods is provided in Section 3.1.

The third stage involves the training of deep sequence models with Long Short-Term Memory (LSTM) networks. These models are particularly apt for time-series data as they are capable of learning long-term dependencies. We adopt different LSTM configurations and tune the hyperparameters to optimize our models for best identifying activity classes. Section 3.2 provides a description of training LSTM networks. Separate deep learning classifiers were created with diverse augmented datasets.

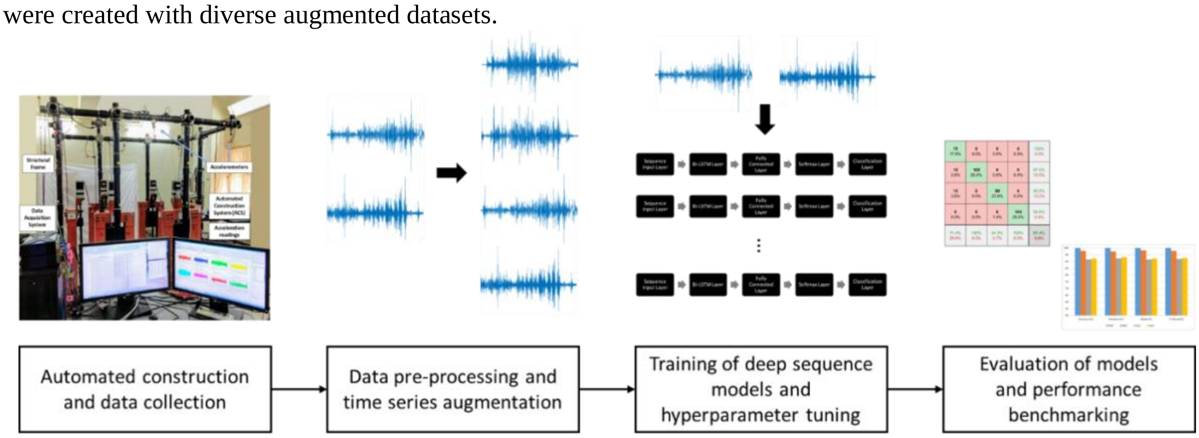

FIG. 1: Overview of the methodology for recognising automated construction activities

Finally, in the fourth stage, we evaluate our trained models to ascertain the best classifier for automated construction monitoring. Evaluation metrics such as accuracy, precision, recall, and F1 score are calculated. Additionally, we assess the influence of the various augmentation techniques on activity classification. As a form of benchmarking, the performance of our LSTM models is compared with that of conventional machine learning classifiers. This comparative evaluation provides a deeper understanding of the performance and suitability of different classifiers in the context of automated construction activity recognition.

3.1. Methods for time series data augmentation

Deep learning methods are known to deliver the best results when there is abundant data available for training. Training deep neural networks to detect common objects may not be changeling since numerous datasets of these objects are publicly available. Creating large datasets of specific objects like common construction equipment (excavators, dump trucks, tower cranes etc.) may be a little more challenging. The data shortage is addressed by generating new data from the existing data by means of various augmentation techniques. Flipping, rotation, cropping etc. are some of the widely used data augmentation methods for image data. The newly generated images create significant variations in the original datasets without altering the original labels. However, augmenting time series data may not be as intuitive as in the case of image datasets (Rashid and Louis, 2019). The newly generated signals should not vary the fundamental characteristics of the original signal in such a way that it may alter the original label. The kinds of variability that create new signals that retain the original labels are random noise, execution method and data collection. The current study introduces these variabilities by jittering, scaling, downsampling and oversampling of the sensor data from the structure.

Jittering: The variability in time series data due to additive sensor noise is introduced through jittering (Rashid and Louis, 2019). White Gaussian noise is incorporated into the raw data to create the jittered dataset (Stahel and Maechler, 2021). The amount of noise varies from -SF * DF/5 to SF * DF/5, where DF is the smallest difference between the values of the measured data and SF is a scaling factor. The value of SF adopted in the current study ranges from 2 to 19.

Scaling: The intensity of vibration corresponds to each construction operation changes with variability in its execution; this variability is introduced through scaling. In scaling, the magnitude of the measured data is altered by multiplying the signal by a scalar (Rashid and Louis, 2019). The scalar value for the current study ranges from 0.3 to 2.1.

Downsampling: The measurements for operation identification can be collected at different sampling rates with varying information contents. Downsampling reduces the sampling rate of the measured data by an integer factor. This data augmentation method is used sparingly to retain the necessary information content for classification. Therefore, the reduction factor ranges from two to five in the current study.

Oversampling: Imbalance in training datasets significantly affects the learning process and often results in high misclassifications of minority classes (Rashid and Louis, 2019). Therefore, oversampling is adopted as a measure to balance the distribution of classes in the datasets. The instances of the underrepresented classes were duplicated in the oversampling. This augmentation method is used along with other methods to create balanced datasets.

3.2. Training of LSTM networks

The data collected in the current study is in the form of time series signals. Long Short-Term Memory (LSTM) networks are suitable for classifying these data since they learn to identify long-term dependency between timesteps of a signal (Hochreiter and Schmidhuber, 1996, 1997; Hochreiter, 1998; Arras *et al.*, 2019). LSTM networks belong to the class of Recurrent Neural Networks (RNN). The architecture of an LSTM network for a classification problem consists of five layers. The first layer is a sequence input layer that inputs the raw sequence data into the network. The second layer is an LSTM layer which learns the long-term dependency between timesteps of the input data. The last three layers, namely, the fully connected layer; SoftMax layer; and classification layer enable the network to predict the class labels. The LSTM layer consists of several LSTM blocks; the flow of information through a block is illustrated in Fig. 2 (Hochreiter and Schmidhuber, 1997; MATLAB & Simulink, 2021). Hidden state (h_t) and cell state (c_t) constitute the state of the layer at timestep t, and x_t denotes the value of the time series at timestep t. The hidden states and cell states are controlled by components such as input gate (i), forget gate (f), cell candidate (g), and output gate (o). The update and reset of the cell state are controlled by the input gate and the forget gate, respectively. The information to the cell state is added by the cell candidate, while information from the cell state to the hidden state is controlled by the output gate. Each of these components can be computed as given in (1) to (4).

$$i_t = \sigma_g(W_i\, x_t + R_i\, h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma_g(W_f\, x_t + R_f\, h_{t-1} + b_f) \quad (2)$$

$$g_t = \sigma_c(W_g\, x_t + R_g\, h_{t-1} + b_g) \quad (3)$$

$$o_t = \sigma_g(W_o\, x_t + R_0\, h_{t-1} + b_o) \quad (4)$$

where W, R and b denote the concatenation of the matrices of the learnable weights such as input weights, recurrent weights and bias of all the components (*i,f,g,o*). The state activation function and gate activation function are represented by σ_c and σ_g. In the current study, x_t represents the acceleration measurements from the structure during automated construction. The raw acceleration data in the form of time series signals were supplied as input to the LSTM network for training. The trained model is evaluated using a separate test dataset.

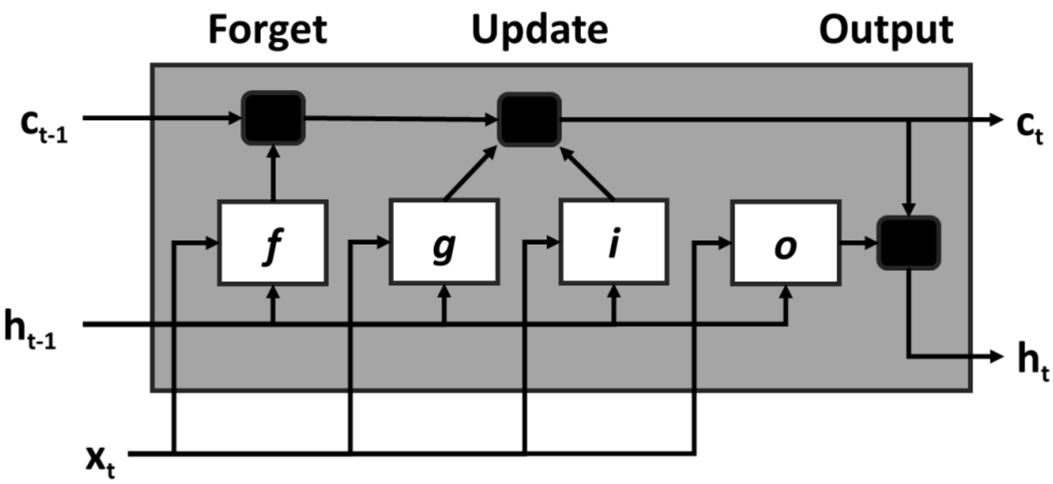

FIG. 2: Information flow in an LSTM block (Hochreiter and Schmidhuber, 1997)

4. CASE STUDY: RECOGNISING AUTOMATED CONSTRUCTION ACTIVITIES

4.1. Automated construction

The automated top-down construction method is used in the ACS prototype deployed in this study (Harichandran *et al.*, 2019b, 2019a, 2020b, 2020a). In terms of the location of the main operating unit and construction scheme, this method is similar to the 'ground factory and building push-up' construction method used for high-rise buildings (Bock and Linner, 2016). The current method, on the other hand, is used to construct the structural frame of low-rise buildings and employs light construction equipment. The main operation and control unit of the ACS is on the ground floor. The structure is built module by module and lifted progressively by the ACS. The upper floors are built first, followed by the lower floors. The platforms or supports at each column location support the structural frame.

Fig. 3 depicts a simplified diagram of automated top-down construction operations. The structural frame is shown in black, while the ACS supports are shown in blue. Only some portions of the structure and the ACS are included in the illustration for clarity. Each column of the structure is made up of several modules that are assembled during construction at each lift level. All operations beginning at a specific lift level are part of the same construction stage. The construction of one floor of a structural frame consists of several construction stages. Structural stability during top-down construction is guaranteed by specially designed configurations and additional supports of the structure. The following are the major operations in one cycle of automated top-down construction: (Fig. 3).

(a) Assemble beam modules and highest column modules on the supports at ground level

(b) Simultaneously lift all supports of the ACS to raise the structure to lift level 1 (Coordinated lifting)

(c) Lower the first support to level 0 while the structure is carried by the remaining supports (Lowering support)

(d) Connect a column module to the unsupported column (Connection of column module)

(e) Lift the support until the load is transferred from the structure (Lifting support)

(f) Repeat steps (c) to (e)

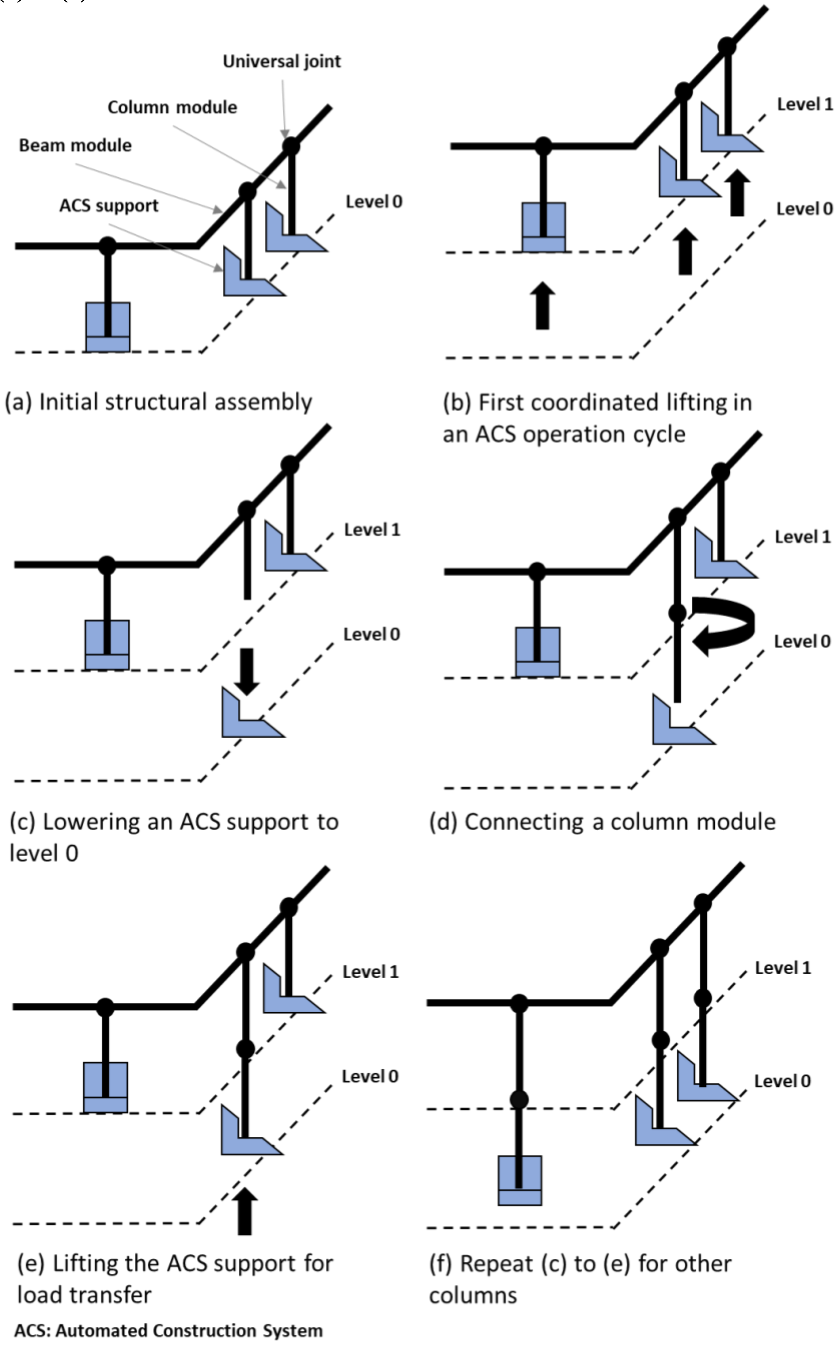

FIG. 3: Schematic representation of automated top-down construction method

After all operations at the first construction stage are completed, the next construction cycle for the second construction stage begins. The current example depicts automated top-down construction with a structure containing redundant supports. Refer to (Harichandran *et al.*, 2020b) for more information on other automated top-down construction schemes for low-rise building construction.

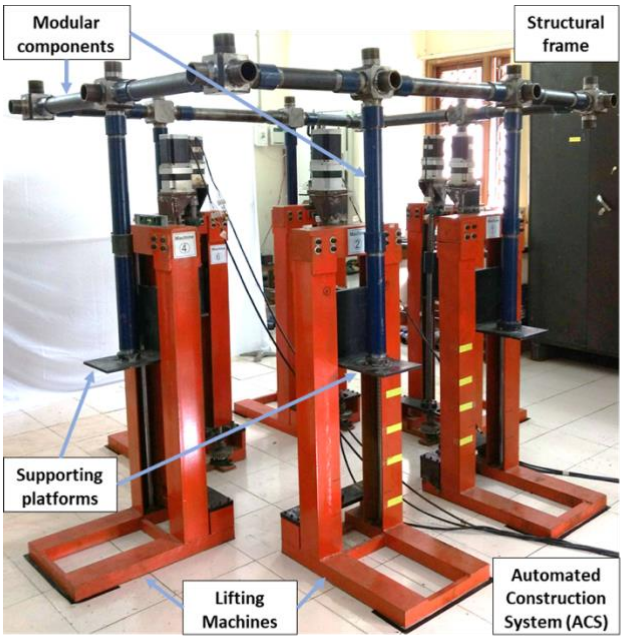

FIG. 4: Partially constructed structural frame on the Automated Construction System (before instrumentation)

4.2. Experiments and data collection

An ACS prototype is deployed to conduct the automated top-down construction in a controlled laboratory environment (Fig. 4). Each experiment consists of two top-down construction cycles that finish two stages of construction. The automated construction experiments were conducted six times. The experimental setup is shown in Fig. 5. Six machines, each with a lifting capacity of two tons make up the ACS prototype. Each column of the structural frame is supported by a platform on the machine. These machines can be arranged at the ground level in adequate numbers depending on the configuration of the structure. All construction-related tasks were automated, except for the connections.

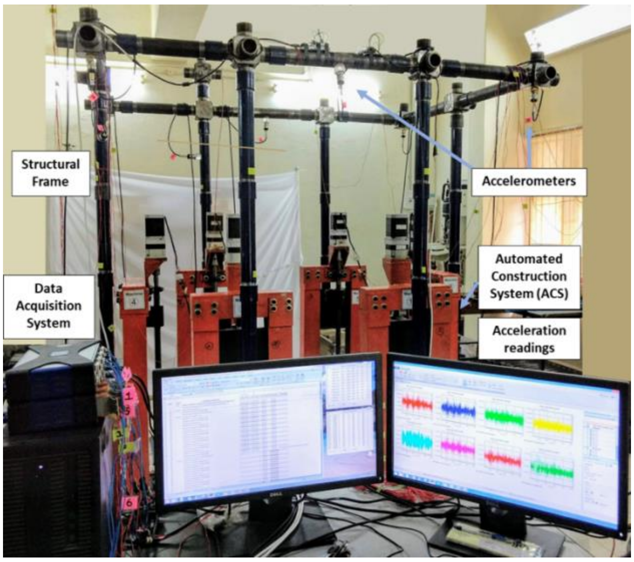

FIG. 5: Experimental setup for monitoring ACS activities

The structural frame in this study is designed to facilitate automated construction. The frame is composed of several small modules that are made from standard steel tube sections. 50 mm nominal bore and 4.5 mm thickness characterize the tube sections. External threading is provided at the ends of every module. Depending on the location, every module is joined to every other module either by a single coupler or by a combination of couplers and universal joints. Couplers are 50 mm nominal bore by 65 mm length steel sockets. The same materials used in the modules are used to make custom steel joints known as universal joints. The universal joints enable the connection of modules in every axial direction.

According to earlier studies in equipment activity recognition and the characteristics of ACS operations, vibration data appear to be the best option for activity recognition. As a result, the accelerometer is chosen to collect data. The placements of the sensors on the structural frame are shown in Fig. 6. The locations are chosen to take into account the practical requirements of ongoing construction, such as seamless data collection, coverage for the entire operation cycle, and high levels of measurable vibrations (high signal to noise ratio). At this stage of the project, more sophisticated sensor placement techniques have not been taken into consideration. Numbers for accelerometers are AM1, AM2, ..., AM8. Data from the sensor is gathered using the HBM universal measuring amplifier (model: QuantumX MX840B). Based on prior research on identifying equipment operation (Ahn *et al.*, 2015; Akhavian and Behzadan, 2015; H. Kim *et al.*, 2018) and the Nyquist criterion (Lyons *et al.*, 2005), the sampling frequency is chosen as 200 Hz. This sampling frequency effectively captured the key operational characteristics. The data can be collected with a timestamp using the HBM data acquisition software Catman (HBM 2020), and it can be displayed right at the moment of recording (Fig. 5). A separate Microsoft Excel file with macro support is kept for manually recording the activities and their duration.

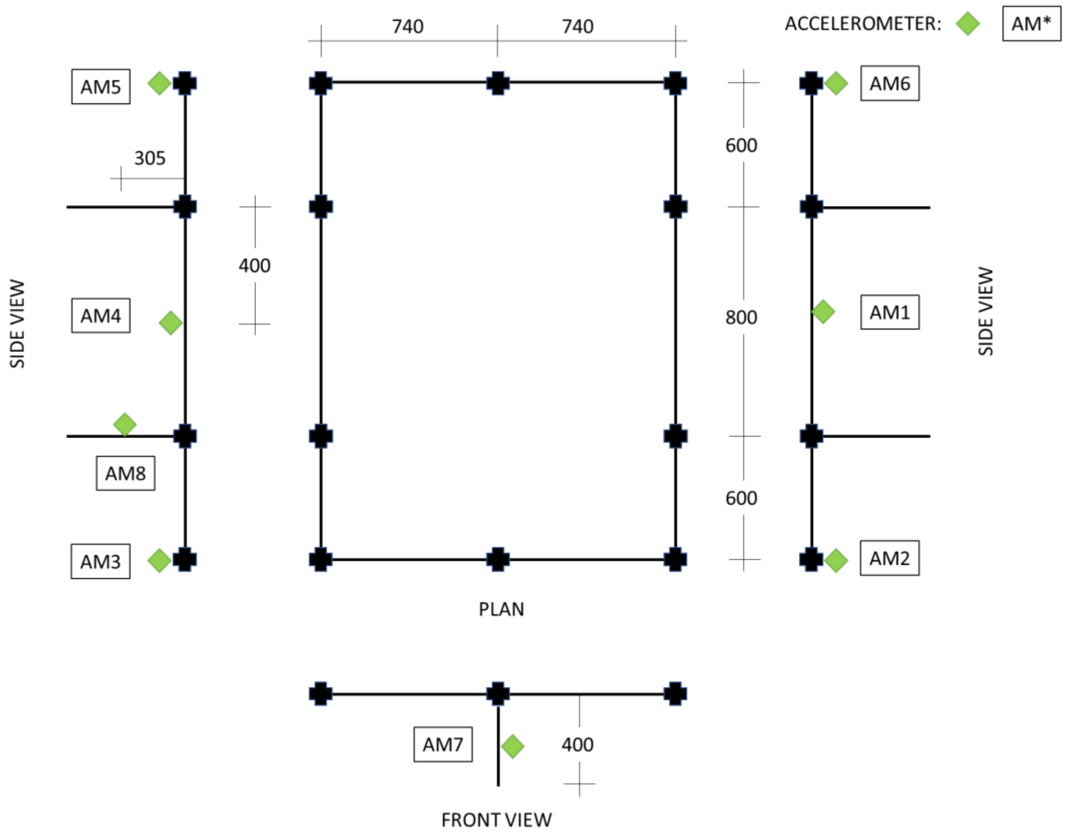

FIG. 6: Sensor locations on the structural frame (All dimensions are in mm)

Fig. 7 shows an acceleration measurement from a single set of experiments. The main operations or states are coloured differently to show the patterns. The *Coordinated Lifting* (shown in orange) appears to have a distinct repetitive pattern. Other operations, on the other hand, appear to have a different pattern at each repetition and construction stage. Consider the blue-coloured *Connection of Column Module* for example.

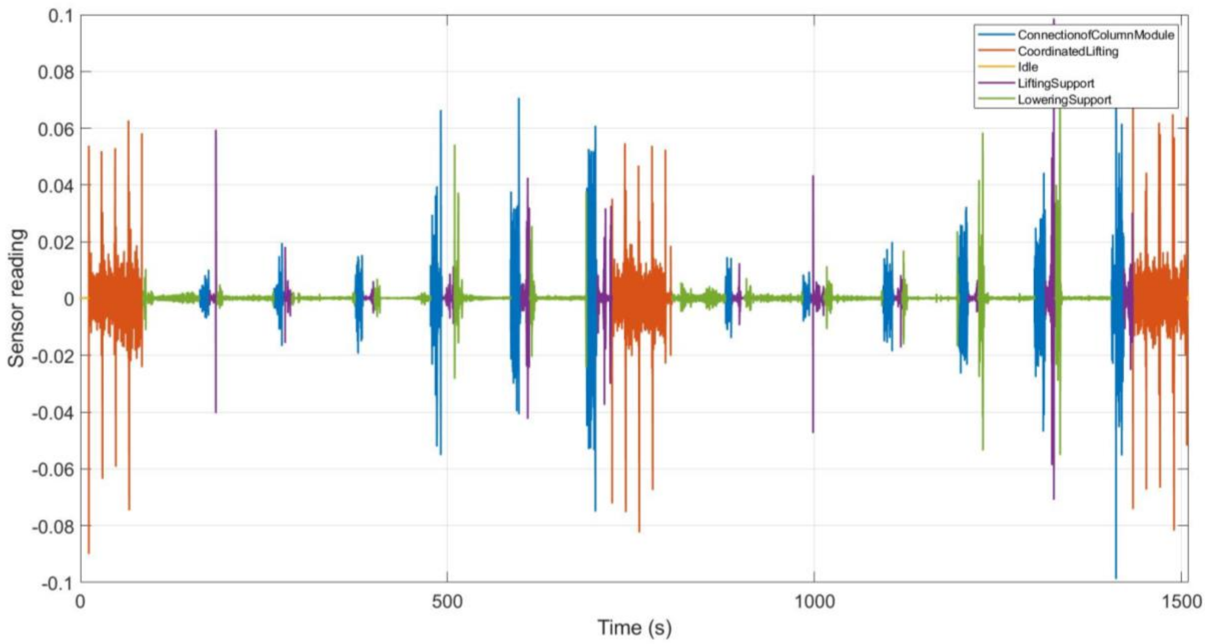

FIG. 7: Acceleration data (sensor: AM4, unit: g) for two cycles of automated top-down construction

4.3. Data pre-processing and time series data augmentation

By means of Catman software (HBM, 2020), the sensor data collected during experiments were imported into Microsoft Excel files and MATLAB files. Further post-processing and analysis of data were carried out in these file formats. The digital records of activities were compared with Catman visualisations and timestamps for creating ground truth labels. Each instance class *(Coordinated Lifting, Lowering Support, Connection of Column Module, Or Lifting Support)* in the raw dataset is split into three in the ratio 80:10:10 for training, validation and testing. Each set is augmented to generate larger datasets through jittering, scaling down sampling and oversampling. The original data points in the *Coordinated Lifting* class are slightly lesser than that of the other classes. Each augmentation method is combined with oversampling to balance the number of data points across the classes. This ensured unbiased and better learning. The raw accelerometer measurements taken from eight locations on the structure were supplied as input data. Since the duration of each operation varies, the length of the input signal also varies. In order to reduce the overall training time, the raw data is truncated to a maximum size of 500 timesteps with a moving window size of 100 timesteps considering the windows having the highest signal magnitude. The training set consists of 4176 data points (1044 instances per class), and the validation set and test set comprise 840 data points each (210 instances per class).

4.4. Training of deep sequence models and hyperparameter tuning

Bidirectional LSTM networks were deployed for identifying automated construction activities. The training data comprises raw acceleration measurements from eight accelerometers to constitute the eight dimensions of the input dataset. The output of the classifiers is the labels of the construction activities. Four classifiers (DL1, DL2, DL3 and DL4) each with different data augmentation techniques were used for this study as given in Table 1. The performance of each classifier is optimized by tuning the hyperparameters on the validation dataset. Several trials of classification experiments were conducted, initially with Bayesian optimization on a wide range of parameters followed by an exhaustive sweep on the selected range. The hyperparameters include initial learning rate, minibatch size and maximum epochs. The hyperparameters of the optimal classifiers are summarized in Table 1. The LSTM networks were trained and evaluated in MATLAB R2020b in a desktop computer with a GPU environment (Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz, 16 GB RAM, Windows 10, 64 bit, GPU: NVIDIA GeForce RTX 2060).

TABLE. 1: *Data augmentation methods and hyperparameters of the deep learning classifiers*

Classifier	Augmentation methods	Learning algorithm	Number of hidden units	Number of iterations	Minibatch size	Learning rate
DL1	Jittering, Oversampling	Bi-LSTM	100	1392	24	0.0009
DL2	Scaling, Oversampling	Bi-LSTM	100	895	28	0.007
DL3	Downsampling, Oversampling	Bi-LSTM	100	2088	18	0.0037
DL4	Jittering, Scaling, Downsampling and Oversampling	Bi-LSTM	100	1566	24	0.0008

4.5. Evaluation of models

The performance of the classifiers was evaluated on the test dataset. Accuracy is selected as the performance index for comparing the LSTM classifiers and comparison with other learning algorithms. The classifiers were also compared by additional performance indices such as precision, recall and F1 score. The equations for these performance indices are given by (5) to (8) where true positive, false positive and false negative are denoted by T_P, F_P and F_N.

$$Accuracy = \frac{Number\ of\ datapoints\ correctly\ recognised}{Total\ number\ of\ datapoints}\ x\ 100\ \% \tag{5}$$

$$Precision = \frac{T_P}{T_P + F_P}\ x\ 100\ \% \tag{6}$$

$$Recall\ = \frac{T_P}{T_P + F_N}\ x\ 100\ \% \tag{7}$$

$$F1\ score = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall}\ x\ 100\ \% \tag{8}$$

In addition to four deep learning classifiers, six machine learning classifiers such as k-Nearest Neighbour (kNN), Decision Tree (DT), Support Vector Machines (SVM), Discriminant Analysis (DA), Naïve Bayes (NB), and Artificial Neural Network (ANN) were deployed for identifying the automated construction activities. Five time domain features (mean, root mean square error, variance, interquartile range and peak values) and five frequency domain features (three of the first main frequencies from Fast Fourier Transform (FFT), period and energy of signal) were extracted for training the machine learning classifiers. The activity recognition performance of the deep sequence models was compared with the traditional machine learning classifiers.

5. RESULTS AND DISCUSSION

5.1. Overview of the results

The deep learning classifiers are trained using different types of augmented datasets and are named DL1, DL2, …, and DL4 as given in Table 1. All of these classifiers are of Bi-LSTM networks which learn sequential information from both ends of the time series. The four main automated construction activities viz. *Connection of column module, Coordinated lifting, Lifting support* and *Lowering support* were identified. The optimal minibatch size of the classifiers ranges from 18 to 28 and the learning rate ranges from 0.0008 to 0.007. The analysis results were summarised in Table 2 and Table 3. Classifiers DL1 to DL3 use datasets predominantly generated by a single augmentation method. While DL4 is trained by a dataset generated as a result of all four data augmentation methods. The machine learning classifiers are numbered from ML1 to ML6, each uses a specific learning algorithm. The time taken for training and hyperparameter tuning of the deep sequence models ranges from more than 8.5 hours to 13.5 hours. However, the total execution time for machine learning models is in seconds. Among the deep learning classifiers, DL3 trained on a down-sampled dataset delivers the best performance with 92.14% accuracy and 92.64 % F1 score, while traditional machine learning classifier ML6 based on ANN secures 100 % accuracy and F1 score.

TABLE. 2: Results of activity recognition by deep learning classifiers

DL Classifier	Augmentation methods	Time for training and hyperparameter tuning (hrs)	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
DL1	Jittering, Oversampling	More than 8.5	91.43	92.23	91.43	91.83
DL2	Scaling, Oversampling	More than 10	77.86	81.97	77.86	79.86
DL3	Downsampling, Oversampling	More than 12	92.14	93.15	92.14	92.64
DL4	Jittering, Scaling, Downsampling and Oversampling	More than 13.5	84.76	85.37	84.76	85.07

TABLE. 3: Results of activity recognition by machine learning classifiers

ML Classifier	Learning algorithm	Total execution time (s)	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
ML1	kNN	1.44	93.16	94.55	94.44	94.50
ML2	DT	0.68	94.87	94.78	94.79	94.79
ML3	SVM	1.56	97.86	97.31	98.26	97.78
ML4	DA	0.64	96.58	97.41	97.22	97.31
ML5	NB	0.70	92.74	94.23	92.01	93.11
ML6	ANN	4.61	100.00	100.00	100.00	100.00

Notes: kNN = k-Nearest Neighbour; DT = Decision Tree; SVM = Support Vector Machines;

DA = Discriminant Analysis; NB = Naïve Bayes; ANN = Artificial Neural Network.

FIG. 8: Performance matrices of machine learning classifiers

5.2. Performance of Machine Learning classifiers vs. Deep Learning classifiers

All of the machine learning classifiers in this study perform at par with or better than deep learning classifiers (LSTM networks) (Fig. 8). In addition to five time domain features, five frequency domain features were also used for training these classifiers. LSTM networks can learn the statistical features from the time series data. However, it cannot directly learn the frequency domain features from the input data. Activity recognition problems that involve operations with signature vibration or frequency require classifiers trained with frequency domain features. LSTM networks trained with limited time domain data will be inadequate for these problems.

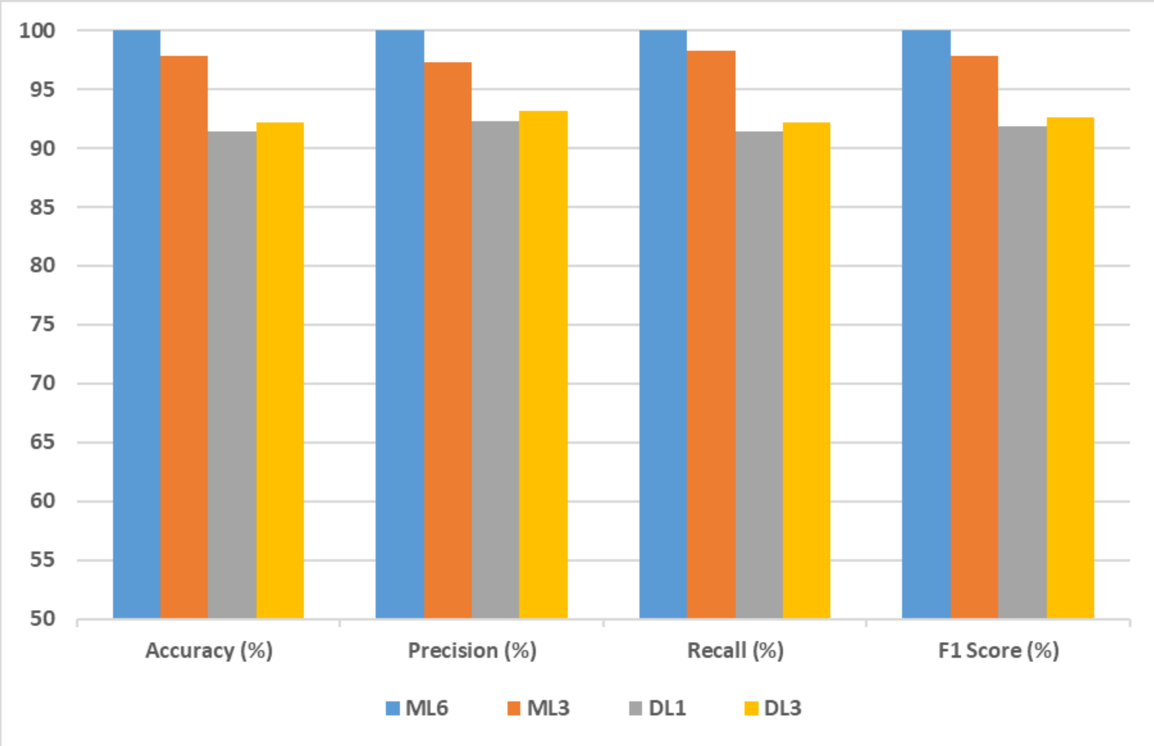

FIG. 9: Comparison of the best two classifiers from machine learning and deep learning classification

Several recently published literature seem to advocate the adaptation of complex deep learning techniques over traditional machine learning algorithms regardless of the application context. Deep learning techniques may deliver superior performance in activity recognition under certain conditions. However, selecting a specific technique for activity recognition has to be made carefully after considering several vital factors. The availability of a large quantum of training data is one of the essential prerequisites. Data augmentation methods are often used to overcome this issue. However, the lack of variety in the original dataset may greatly affect the performance of the classifier. This may often lead to overfitting. In this scenario, conventional machine learning methods offer a better solution. Consider Fig. 9 which compares the best two classifiers from machine learning and deep learning classification. The best-performing classifier ML6 is an artificial neural network with a simple architecture. The second-best classifier is an SVM with a polynomial kernel. None of the deep learning classifiers presented in this study could match the performance of the conventional machine learning classifiers.

Some of the previous studies used complex or hybrid deep learning classifiers for construction equipment activity recognition. The comparison of activity identification performances of these methods and that of the current study is given in Table 4. These results show that the complexity of the learning algorithm may not always ensure better performance. The time, cost and effort in collecting good quality data and developing a complex classifier have to be justified by its performance in the application. Ensuring the best performance of deep learning algorithms requires in-depth knowledge of the network architecture. The scarcity of in-house deep learning experts in the construction engineering domain makes the actual implementation of these techniques expensive (Akinosho *et al.*, 2020).

TABLE. 4: Comparison of activity recognition performance with previous studies

Reference	Equipment	Data collected	Number of activity classes	Methods/Algorithms	Performance in activity identification
(Luo *et al.*, 2018)	Workers and various equipment	Images	17	Faster R-CNN + ResNet-50, Relevance Networks	Precision: 62.4% Recall: 87.3%.
(Kim *et al.*, 2019)	Excavator	Images	4	TLD, Hybrid network (CNN and LSTM)	Average accuracy: 93.8%
(Roberts and Golparvar-Fard, 2019)	Excavator and dump truck	Videos	Excavator: 5 Dump truck: 2	CNN (ResNeXt-101), HMM, GMM, SVM	Accuracy: for excavators 86.8% for dump trucks 88.5%
(Rashid and Louis, 2019)	Excavator and front-end loader	IMU data	Excavator: 9 Front-end loader: 10	LSTM	Accuracy: for excavator 97.9% for front-end loader 96.7% F1 score: for excavator 97.6% for front-end loader 96.3%
(Chen *et al.*, 2020)	Excavators	Videos	3	Faster R-CNN	Overall accuracy 87.6%
(Slaton *et al.*, 2020)	Roller compactor and excavator	Acceleration	Roller compactor: 6 Excavator: 7	Hybrid network (CNN and LSTM)	Accuracy: for compactor: 77.1% for excavator: 77.6%
(Scarpiniti *et al.*, 2021)	10 construction equipment	Sound	10	DBN	Overall performance Accuracy: 97.79% Precision: 97.80% Recall: 97.79% F1 score: 97.79%
(Sherafat *et al.*, 2022)	Excavator and loader	Sound	Excavator: 4 Loader: 3	CNN	Accuracy: for excavator:88.4% for loader: 87.1%
(Kim *et al.*, 2023)	Excavator	Videos	4	CNN(GoogleNet) and Bi-LSTM	Accuracy: 87.5%
(Chen *et al.*, 2023)	Excavator and loader	Videos	Excavator: 2 Loader: 2	YoloV5, SORT, and CLIP	Accuracy: for excavator: 86% for loader: 82.5%
Current study	ACS	Acceleration	4	Bi-LSTM	Accuracy: 92.14% Precision: 93.15% Recall: 92.14% F1 score: 92.64%
Current study	ACS	Acceleration	4	ANN	Accuracy: 100% Precision: 100% Recall: 100% F1 score: 100%

Notes: CNN = Convolutional Neural Network; TDL= Tracking-Learning-Detection; LSTM = Long Short Term Memory Network; HMM = Hidden Markov Model; GMM = Gaussian Mixture Model; SVM = Support Vector Machines; DBN: Deep Belief Network; YoloV5= You Look Only Once; SORT= Simple Online and Real-Time Tracking; CLIP= Contrastive Language Image Pre-training; ACS = Automated Construction System; ANN = Artificial Neural Network.

5.3. Influence of data augmentation methods on activity recognition

Without the additional cost of collecting a large quantum of data or augmentation or huge training time the conventional machine learning classifiers delivered better results. The current study also shows the capability of conventional machine learning classifiers in activity recognition for a sparsely explored application domain. The need to select relevant hand-crafted features is considered one of the major drawbacks of conventional machine learning techniques. The introduction of deep learning may avoid the feature selection step. However, domain knowledge is necessary to develop a robust classifier for activity recognition. Consider Fig. 10 which shows the influence of the data augmentation techniques on the performance of the classifiers. All of the augmentation methods except scaling result in classifiers with more than 80% accuracy as well as F1 score. While using scaling for data augmentation, the accuracy drops as low as 77.86%. Intuitively, the introduction of a variety of data augmentation methods should enhance the performance of a classifier. However, the classifier DL4 which uses all of the data augmentation methods is one of the worst-performing classifiers. This is mainly because of the presence of the scaled dataset. Hence, it is evident that the amplitude of the time series data plays a significant role in distinguishing the classes. This can be also confirmed by visual inspection of the measured data (Fig. 7). Therefore, the data augmentation methods have to be carefully selected in such a way that they should not affect the characteristics of the original dataset.

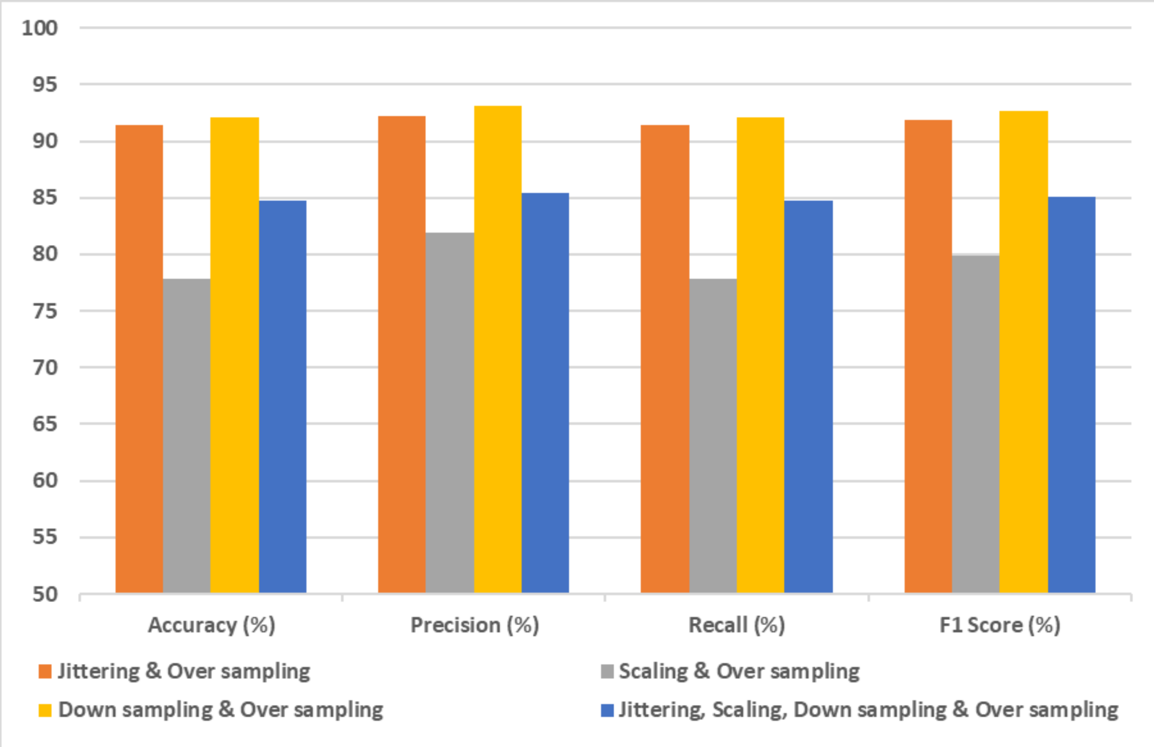

FIG. 10: Effect of data augmentation techniques on time series classification

5.4. Performance comparison with confusion matrices

Performance measures such as accuracy, F1 score, precision and recall provide an overall performance of the classifiers. However detailed information on misclassified activities is obtained through confusion matrices Fig. 11 to Fig. 14 depict the confusion matrices of the LSTM classifiers. The rows represent the output classes or the predicted classes, while the columns represent target classes or actual classes. The elements in the main diagonal are correctly classified and off-diagonal elements are misclassified. Each cell in the matrix contains the number of instances and the percentage of the total number of instances. The last column on the plot shows the percentage of instances predicted in each class that are correctly classified, termed as precision and incorrectly classified, termed as false discovery rate. The last row on the plot shows the percentage of correctly classified instances in each class termed as recall and incorrectly classified instances termed as false negative rate. The last cell of the matrix (bottom right) shows the overall accuracy.

The DL1 classifier that majorly uses a jittered data set has high precision and recall for all classes, except *Connection of Column Module*. Fifteen instances each were misclassified as *Coordinated Lifting* and *Lifting Support*. These misclassifications contribute the most to the total error of 8.6%. The DL2 classifier that prominently uses scaled datasets delivers the lowest performance with 77.9% accuracy. The *Lifting Support* has been misclassified as *Lowering Support* in 48 instances (11.4% of total instances) and *Connection of Column*

Module in 15 instances (3.6%) to result in a low recall of 35.2%. The signals for *Lifting Support* have a characteristic peak towards the end while the support makes contact with the column, whereas the signals for *Lowering Support* tend to have a peak in the beginning while the support detaches from the column (Fig. 7). Augmentation by scaling changes the amplitude of the signal and contribute towards interclass confusion. Therefore, the DL2 classifier makes several misclassifications. The DL3 classifier whose dataset mostly contains down-sampled data has the highest overall accuracy of 92.1%. The majority of the misclassifications (54 instances) are due to the interclass confusion of *Lowering Support* with *Lifting Support.* The DL4 classifier contains all four types of augmented datasets including jittering, scaling, downsampling and oversampling. The *Lifting Support* has the lowest recall value of 62.9% where 60 instances were misidentified as *Lowering Support.* Irrespective of the classifiers, all instances of *Coordinated Lifting* are correctly identified to result in 100% recall. The reason for the result is that *Coordinated Lifting* has a clear repetitive pattern unlike the other classes (Fig. 7).

FIG. 11: Confusion matrix for DL1 using jittering and oversampling

FIG. 12: Confusion matrix for DL2 using scaling and oversampling

FIG. 13: Confusion matrix for DL3 using downsampling and oversampling

FIG. 14: Confusion matrix for DL4 using jittering, scaling, downsampling and oversampling

6. CONCLUSIONS

Recognising activities of construction equipment is essential for monitoring productivity, construction progress, safety and environmental impacts. Advanced deep learning methods are widely applied for automatically identifying equipment activities. The existing studies on construction equipment activity recognition are focused mainly on earth excavation and moving equipment and methods for identifying activities of Automated Construction Systems (ACS) are limited. This study evaluates the suitability of deep sequence models in activity recognition of automated construction equipment through a case study on a low-rise ACS. The activities of an Automated Construction System were identified by deep learning classifiers (LSTM) and their performance was benchmarked with conventional machine learning classifiers (kNN, DT, SVM, DA, NB, and ANN). Diverse augmentation methods were adopted for generating datasets for training the deep learning classifiers. Regardless of the application context, several recently published literature appear to support the adoption of sophisticated deep learning techniques. However, the results of activity identification in the current study show that all of the conventional machine learning classifiers perform equivalently or better than deep learning classifiers.

The classifiers for equipment activity recognition must be selected based on the identification problem and availability of datasets. Implementation of traditional machine learning for construction activity recognition is more feasible than that of deep learning. The actual implementation of deep learning methods in the construction industry demands high investment in terms of time, cost, and effort to collect good quality data in addition to high training time and computational power. In contrast, simple machine learning algorithms with hand-crafted features may offer better performance compared to complex algorithms. Complex learning algorithms need not necessarily result in better performance. The lack of variety in the original dataset during data augmentation greatly affects the performance of the deep learning classifiers. Therefore, the traditional machine learning classifiers outperformed the deep learning classifiers in identifying automated construction activities. Besides, augmenting data may not deliver better results if it alters the characteristics of the original dataset. The selection of appropriate data augmentation methods and the design of network architecture demands great expertise. Even though deep learning may avoid feature selection, domain knowledge is necessary to develop a robust classifier for activity recognition. In this study, machine learning classifiers trained with frequency-domain features delivered better results in identifying activities with signature vibration or frequency. LSTM classifiers trained with limited time-domain data seem to be inadequate for vibration-based activity recognition.

This study presents noteworthy contributions to the field of construction automation and activity recognition, advancing the understanding of how computational methods can be utilized to formalize complex engineering processes. It examines the feasibility of deploying deep sequence models for developing an activity recognition framework for ACS, an area that has been under-researched. The novelty of this study is further emphasized by implementing and testing machine learning and deep learning algorithms on a real-world scale ACS, one of the first instances in the emerging field of industrialized construction. Contrary to widely-held beliefs about the superiority of deep learning, this study presents counterintuitive findings that conventional machine learning algorithms can deliver equivalent or superior performance in certain contexts. Such insights enrich our understanding of the comparative efficiencies of these computational methods in codifying complex construction processes. The study also elucidates the impact of data augmentation methods on deep learning classifiers' performance in recognizing construction activities, shedding light on the need for judicious selection and application of these methods. This reinforces the need for careful customization of computational tools to effectively encapsulate complex engineering knowledge. The findings of this study contribute towards developing an integrated monitoring framework for low-rise ACS, a feature that is mostly only associated with high-rise ACS due to their complexity. In essence, this study provides a significant contribution by enhancing our understanding of the practical applications and optimization of machine learning and deep learning techniques within construction automation and activity recognition.

The findings in this study provide a significant revelation that conventional ML algorithms can achieve comparable, if not superior, performance to deep learning methods in the context of low-rise ACS. These findings carry crucial implications for the construction industry. Implementation of traditional ML methods instead of deep learning allows for cost-effective, accessible, and streamlined adoption. Unlike deep learning, which often requires extensive data and specialized expertise, conventional ML provides flexibility and ease of implementation. This aligns not only with the practical and budgetary concerns of several stakeholders but also with broader sustainability goals, due to reduced energy consumption. Furthermore, it provides strategic flexibility to decision-makers by allowing them to choose between computational methods based on their specific needs and goals. The findings also support a more pragmatic approach to construction automation, emphasizing opportunities for increased accessibility, customization, and alignment with the unique requirements of the construction environment.

7. LIMITATIONS AND FUTURE WORK

The automated construction experiments in this study are conducted in a controlled laboratory environment. Vibration measurements from the ACS while implemented in a construction site may encounter more ambient disturbances and complex operating scenarios. In addition to that, finer activity classification is required for estimating the productivity of the equipment. Future work focuses on identifying the low-level activities of the ACS. The possibility of incorporating domain knowledge such as activity sequence and hierarchical relationships in activity recognition is also under consideration.

ACKNOWLEDGEMENTS

Support from staff and faculty members of the Building Automation Laboratory, Structural Engineering Laboratory and Civil Engineering workshop of Indian Institute of Technology Madras is greatly acknowledged. This project is funded by the Department of Science and Technology (DST) and Science and Engineering Research Board (SERB), India, through the grants DST/TSG/AMT/2015/234 and IMP/2018/000224.

REFERENCES

Ahn, C. R., Lee, S., Peña, F., et al. (2015) 'Application of Low-Cost Accelerometers for Measuring the Operational Efficiency of a Construction Equipment Fleet', Journal of Computing in Civil Engineering, 29(2), p. 04014042. doi: 10.1061/(ASCE)CP.1943.

Akhavian, R. and Behzadan, A. H. (2015) 'Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers', Advanced Engineering Informatics, 29, pp. 867–877. doi: 10.1016/j.aei.2015.03.001.

Akinosho, T. D., Oyedele, L. O., Bilal, M., et al. (2020) 'Deep Learning in the Construction Industry: A Review of Present Status and Future Innovations', Journal of Building Engineering, p. 101827. doi: 10.1016/j.jobe.2020.101827.

Arras, L., Arjona-Medina, J., Widrich, M., et al. (2019) 'Explaining and Interpreting LSTMs', in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 211–238. doi: 10.1007/978-3-030-28954-6_11.

Bock, T. and Linner, T. (2016) Site Automation Automated/Robotic On-site Factories. 1st edn. New York: Cambridge University Press.

Bureau of Labor Statistics (2021) National census of fatal occupational injuries in 2021. Available at: www.bls.gov/iif.

Cai, S., Ma, Z., Skibniewski, M. J., et al. (2019) 'Construction automation and robotics for high-rise buildings over the past decades: A comprehensive review', Advanced Engineering Informatics, 42, p. 100989. doi: 10.1016/J.AEI.2019.100989.

Cao, J., Huang, W., Zhao, T., et al. (2017) 'An enhance excavation equipments classification algorithm based on acoustic spectrum dynamic feature', Multidimensional Systems and Signal Processing, 28(3), pp. 921–943. doi: 10.1007/s11045-015-0374-z.

Cao, J., Wang, W., Wang, J., et al. (2017) 'Excavation Equipment Recognition Based on Novel Acoustic Statistical Features', IEEE transactions on cybernetics, 47(12), pp. 4392–4404. doi: 10.1109/TCYB.2016.2609999.

Castro-Lacouture, D. (2009) 'Construction Automation', in Nof, S. Y. (ed.) Springer Handbook of Automation. Springer, pp. 1063–1078.

Chen, C., Xiao, B., Zhang, Y., et al. (2023) 'Automatic vision-based calculation of excavator earthmoving productivity using zero-shot learning activity recognition', Automation in Construction, 146, p. 104702. doi: 10.1016/j.autcon.2022.104702.

Chen, C., Zhu, Z. and Hammad, A. (2020) 'Automated excavators activity recognition and productivity analysis from construction site surveillance videos', Automation in Construction, 110, p. 103045. doi: 10.1016/j.autcon.2019.103045.

Cheng, C.-F., Rashidi, A., Davenport, M. A., et al. (2017) 'Activity analysis of construction equipment using audio signals and support vector machines', Automation in Construction, 81, pp. 240–253. doi: 10.1016/J.AUTCON.2017.06.005.

Forestier, G., Petitjean, F., Dau, H. A., et al. (2017) 'Generating Synthetic Time Series to Augment Sparse Datasets - IEEE Conference Publication', in IEEE International Conference on Data Mining (ICDM). New Orleans, LA, USA, pp. 865–870. Available at: https://ieeexplore.ieee.org/document/8215569.

Gassel, F. van (2005) 'The Development of a Concept for a Dutch Construction System for High-Rise Buildings', in 22nd International Symposium on Automation and Robotics in Construction (ISARC 2005). Ferrara,Italy, pp. 1–4. doi: 10.22260/ISARC2005/0042.

Ghelmani, A. and Hammad, A. (2023) 'Self-supervised contrastive video representation learning for construction equipment activity recognition on limited dataset', Automation in Construction, 154, p. 105001. doi: 10.1016/J.AUTCON.2023.105001.

Golparvar-Fard, M., Heydarian, A. and Niebles, J. C. (2013) 'Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers', Advanced Engineering Informatics, 27(4), pp. 652–663. doi: 10.1016/J.AEI.2013.09.001.

Hamada, K., Furuya, N., Inoue, Y., et al. (1998) 'Development of automated construction system for high-rise reinforced concrete buildings', in IEEE International Conference on Robotics and Automation. Leuven, Belgium: IEEE, pp. 2428–2433. doi: 10.1109/ROBOT.1998.680704.

Harichandran, A., Raphael, B. and Mukherjee, A. (2019a) 'Determination of Automated Construction Operations from Sensor Data Using Machine Learning', in 4th International Conference on Civil and Building Engineering Informatics. Sendai, Japan, pp. 77–84.

Harichandran, A., Raphael, B. and Mukherjee, A. (2019b) 'Identification of the Structural State in Automated Modular Construction', in 36th International Symposium on Automation and Robotics in Construction (ISARC 2019). Banff, Canada, pp. 187–193. doi: https://doi.org/10.22260/ISARC2019/0026.

Harichandran, A., Raphael, B. and Mukherjee, A. (2020a) 'A Robust Framework for Identifying Automated Construction Operations', in 37th International Symposium on Automation and Robotics in Construction (ISARC 2020). Kitakyushu, Japan, pp. 473–480. doi: https://doi.org/10.22260/ISARC2020/0066.

Harichandran, A., Raphael, B. and Mukherjee, A. (2020b) 'Development of Automated Top-Down Construction System for Low-rise Building Structures', International journal of industrialized construction, 1(1), pp. 22–33. doi: doi.org/10.29173/ijic217.

Harichandran, A., Raphael, B. and Mukherjee, A. (2021) 'A hierarchical machine learning framework for the identification of automated construction operations', Journal of Information Technology in Construction, 26, pp. 591–623. doi: 10.36680/j.itcon.2021.031.

Harichandran, A., Raphael, B. and Mukherjee, A. (2023) 'Equipment activity recognition and early fault detection in automated construction through a hybrid machine learning framework', Computer-Aided Civil and Infrastructure Engineering, 38(2), pp. 253–268. doi: 10.1111/mice.12848.

HBM (2020) catman Data Acquisition Software, HBM. Available at: https://www.hbm.com/en/2290/catman-data-acquisition-software/ (Accessed: 7 September 2020).

Hochreiter, S. (1998) 'The vanishing gradient problem during learning recurrent neural nets and problem solutions', International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems, 6(2), pp. 107–116. doi: 10.1142/S0218488598000094.

Hochreiter, S. and Schmidhuber, J. (1996) 'LSTM can solve hard long time lag problems', in Proceedings of the 9th International Conference on Neural Information Processing Systems (NIPS'96), pp. 473–479. Available at: https://dl.acm.org/doi/10.5555/2998981.2999048.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long Short-Term Memory', Neural Computation, 9(8), pp. 1735–1780. doi: 10.1162/neco.1997.9.8.1735.

Hu, H., Kantardzic, M. and Sethi, T. S. (2020) 'No Free Lunch Theorem for concept drift detection in streaming data classification: A review', Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(2), p. e1327. doi: 10.1002/WIDM.1327.

Kang, T.-K., Nam, C., Lee, U.-K., et al. (2011) 'Development of Robotic-Crane Based Automatic Construction System for Steel Structures of High-Rise Buildings', in 28th International Symposium on Automation and Robotics in Construction (ISARC 2011). Seoul, Korea, pp. 670–671. doi: 10.22260/ISARC2011/0125.

Kim, H., Ahn, C. R., Engelhaupt, D., et al. (2018) 'Application of dynamic time warping to the recognition of mixed equipment activities in cycle time measurement', Automation in Construction, 87, pp. 225–234. doi: 10.1016/J.AUTCON.2017.12.014.

Kim, I. S., Latif, K., Kim, J., et al. (2023) 'Vision-Based Activity Classification of Excavators by Bidirectional LSTM', Applied Sciences (Switzerland), 13(1). doi: 10.3390/APP13010272.

Kim, J. and Chi, S. (2019) 'Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles', Automation in Construction, 104, pp. 255–264. doi: 10.1016/j.autcon.2019.03.025.

Kim, J., Chi, S. and Ahn, C. R. (2021) 'Hybrid kinematic–visual sensing approach for activity recognition of construction equipment', Journal of Building Engineering, 44. doi: 10.1016/j.jobe.2021.102709.

Kim, J., Chi, S. and Choi, M. (2019) 'Sequential Pattern Learning of Visual Features and Operation Cycles for Vision-Based Action Recognition of Earthmoving Excavators', in Computing in Civil Engineering 2019: Data, Sensing, and Analytics - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2019. American Society of Civil Engineers (ASCE), pp. 298–304. doi: 10.1061/9780784482438.038.

Kim, J., Chi, S. and Seo, J. (2018) 'Interaction analysis for vision-based activity identification of earthmoving excavators and dump trucks', Automation in Construction, 87, pp. 297–308. doi: 10.1016/j.autcon.2017.12.016.

Langroodi, A. K., Vahdatikhaki, F. and Doree, A. (2021) 'Activity recognition of construction equipment using fractional random forest', Automation in Construction, 122, p. 103465. doi: 10.1016/j.autcon.2020.103465.

Luo, X., Li, ; Heng, Cao, D., et al. (2018) 'Recognizing Diverse Construction Activities in Site Images via Relevance Networks of Construction-Related Objects Detected by Convolutional Neural Networks'. doi: 10.1061/(ASCE)CP.1943-5487.0000756.

Lyons, G. M., Culhane, K. M., Hilton, D., et al. (2005) 'A description of an accelerometer-based mobility monitoring technique', Medical Engineering & Physics, 27(6), pp. 497–504. doi: 10.1016/J.MEDENGPHY.2004.11.006.

MATLAB & Simulink (2021) Long Short-Term Memory Networks. Available at: https://in.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html#mw_0862a318-9d89-4952-8803-7b529c62214e (Accessed: 5 September 2021).

Meng, Q. and Zhu, S. (2022) 'Construction Activity Classification Based on Vibration Monitoring Data: A Supervised Deep-Learning Approach with Time Series RandAugment', Journal of Construction Engineering and Management, 148(9). doi: 10.1061/(ASCE)CO.1943-7862.0002359.

Raphael, B., Rao, K. S. C. and Varghese, K. (2016) 'Automation of modular assembly of structural frames for buildings', in 33rd International Symposium on Automation and Robotics in Construction (ISARC 2016). Auburn, USA, pp. 412–420. doi: https://doi.org/10.22260/ISARC2016/0050.

Rashid, K. M. and Louis, J. (2019) 'Times-series data augmentation and deep learning for construction equipment activity recognition', Advanced Engineering Informatics, 42, p. 100944. doi: 10.1016/j.aei.2019.100944.

Rashid, K. M. and Louis, J. (2020) 'Activity identification in modular construction using audio signals and machine learning', Automation in Construction, 119, p. 103361. doi: 10.1016/j.autcon.2020.103361.

Roberts, D. and Golparvar-Fard, M. (2019) 'End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level', Automation in Construction, 105, p. 102811. doi: 10.1016/j.autcon.2019.04.006.

Scarpiniti, M., Colasante, F., Di Tanna, S., et al. (2021) 'Deep Belief Network based audio classification for construction sites monitoring', Expert Systems with Applications, 177. doi: 10.1016/J.ESWA.2021.114839.

Sekiguchi, T., Honma, K., Mizutani, R., et al. (1997) 'The Development and Application of an Automatic Building Construction System Using Push-Up Machines', in Proceedings of the 14th International Symposium on Automation and Robotics in Construction (ISARC). Pittsburgh, USA: International Association for Automation and Robotics in Construction (IAARC), pp. 321–328. doi: 10.22260/isarc1997/0040.

Sherafat, B., Ahn, C. R., Akhavian, R., et al. (2020) 'Automated Methods for Activity Recognition of Construction Workers and Equipment: State-of-the-Art Review', Journal of Construction Engineering and Management, 146(6), p. 03120002. doi: 10.1061/(ASCE)CO.1943-7862.0001843.

Sherafat, B., Rashidi, A. and Asgari, S. (2022) 'Sound-based multiple-equipment activity recognition using convolutional neural networks', Automation in Construction, 135, p. 104104. doi: 10.1016/J.AUTCON.2021.104104.

Shi, Y., Xia, Y., Zhang, Y., et al. (2020) 'Intelligent identification for working-cycle stages of excavator based on main pump pressure', Automation in Construction, 109, p. 102991. doi: 10.1016/j.autcon.2019.102991.

Slaton, T., Hernandez, C. and Akhavian, R. (2020) 'Construction activity recognition with convolutional recurrent networks', Automation in Construction, 113, p. 103138. doi: 10.1016/j.autcon.2020.103138.

Stahel, W. and Maechler, M. (2021) R: 'Jitter' (Add Noise) to Numbers. Available at: https://stat.ethz.ch/R-manual/R-patched/library/base/html/jitter.html (Accessed: 19 June 2021).

Wakisaka, T., Furuya, N., Inoue, Y., et al. (2000) 'Automated construction system for high-rise reinforced concrete buildings', Automation in Construction, 9(3), pp. 229–250. doi: 10.1016/S0926-5805(99)00039-4.

Yamazaki, Y. and Maeda, J. (1998) 'The SMART system: an integrated application of automation and information technology in production process', Computers in Industry, 35(1), pp. 87–99. doi: 10.1016/S0166-3615(97)00086-9.