Northumbria Research Link

Citation: Kassem, Mohamad, Mahamedi, Elham, Rogage, Kay, Duffy, Kieren and Huntingdon, James (2021) Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach. Automation in Construction, 124. p. 103532. ISSN 0926-5805

Published by: Elsevier

URL: https://doi.org/10.1016/j.autcon.2020.103532 <https://doi.org/10.1016/j.autcon.2020.103532>

This version was downloaded from Northumbria Research Link: http://nrl.northumbria.ac.uk/id/eprint/45641/

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: http://nrl.northumbria.ac.uk/policies.html

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. and Huntingdon, J. (2021). Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction, Volume 124, 2021, 103532, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103532.

Measuring and benchmarking the productivity of excavators in infrastructure projects: A Deep Neural Network approach

Abstract

Inefficiencies in the management of earthmoving equipment greatly contribute to the productivity gap of infrastructure projects. This paper develops and tests a deep Neural Network (DNN) model for estimating the productivity of excavators and establishing a productivity measure for their benchmark. After investigating current practices for measuring the productivity of earthwork equipment during 13 interviews with selected industry experts, the DNN model was developed and tested in one of the 'High Speed rail second phase' (HS2) sites.

The accuracy of prediction achieved by the DNN model was evaluated using the coefficient of determination (R2) and the Weighted Absolute Percentage Error (WAPE) resulting in 0.87 and 69.64%, respectively. This is an adequate level of accuracy when compared to other similar studies. However, according to the WAPE method, the accuracy is still 10.36% below the threshold (i.e. 80%) expected by industry experts. An inspection of the prediction results over the testing period (21 days) revealed better precision in days with high excavation volumes compared to days with low excavation volumes. This was attributed to the likely involvement of manual work (i.e. archaeologists in the case of the selected site) alongside some of the excavators, which caused gaps in telematics data. This indicates that the accuracy attained is adequate, but the proposed approach is more accurate in a highly mechanised environment (i.e. excavation work with equipment predominantly and limited manual interventions) compared to a mixed mechanised-manual working environment. A bottom-up benchmark measure (i.e. excavation rate) that can be used to measure and benchmark the excavation performance of an individual or a group of equipment, through a work area, to a whole site was also proposed and discussed.

Keywords: Deep Neural Network; Earthwork; Machine Learning; Telematics.

1 Introduction

The need to improve the performance of the construction sector has been a recurrent theme for a long time. The sector has been infamous for inefficient practices (e.g. Latham, 1994; Egan, 1998; Wolstenholme, 2009; Hackitt, 2018; Farmer, 2016; KPMG, 2015; CITB, 2018) and its productivity has increased by just 1% in over 20 years (McKinsey, 2017). Plants and equipment are among the most critical resources to the performance of construction sites. They have a significant impact on the total cost of projects (European Commission, 2014); are a critical 'bottleneck' resource that is often linked to delays (Ok and Sinha, 2006; Kassem et al., 2019), and a major contributor to on- and off- site congestion and air pollution (Greater London Authority, 2014). Hence, their effective management represents a key opportunity for economic, environmental and safety efficiency gains.

Applications of machine learning to address challenges within the construction industry are on the rise. Machine learning techniques are used to understand and predict project delays (Asadi et al., 2015; Gondia et al., 2020); improve design of built assets (Ngo, 2018; Singaravel et al., 2018); manage site safety (Nath et al., 2020; Baker et al., 2020), aid offsite construction (Rashid et al., 2020), classify and identify components for facilities management (Marzouk and Zaher, 2020), among several other applications. This trend is supported by the growing availability of digital data across design, construction, and operation of built assets. However, very few studies investigating the productivity of earthwork in infrastructure projects are available as evidenced by the literature review in Section 2.

The estimation of productivity of equipment (such as excavators) is a key challenge in earthwork on infrastructure project sites. Another challenge is the lack of performance measures for their productivity benchmark that can be automatically computed using machine learning. This paper develops and tests a Deep Neural Network (DNN) model to measure the productivity of excavators based on the volume of earth removed, using an array of telematics data fields as feature inputs. It also proposes a measure for benchmarking the performance of excavators and uses the developed DNN model to calculate it. DNN is a kind of Artificial Neural Network (ANN) having multiple hidden layers present in between the input and output layer (Romdhani, 2015) and can be used to model a complicated non-linear relationship between input and output factors (Dixit et al., 2018).

The paper is organised as follows: Section 2 classifies and reviews existing studies and provides evidence of the research gap; Section 3 presents the relevant findings from the 13 interviews carried out to understand practices and challenges facing the estimation of earthwork productivity; Section 4 demonstrates the steps of developing and optimising the DNN model for estimating the volume of earth removed by excavators; Section 5 performs the testing and evaluation of the DNN model and demonstrate its use in calculating the proposed performance measure (i.e. excavation rate) for benchmarking productivity of excavators; Section 6 discusses the findings and outlines future developments; and finally Section 7 concludes and describes implications.

2 Literature review

The effective management of equipment in infrastructure projects is a critical function for general contractors and specialised earthwork contractors due to their impact on project cost, schedule, overall business profitability, environment (i.e. CO2 emissions and NOx levels), and health and safety. Current practices in management of site equipment are still highly dependent on the site managers' experience . The availability of data relating to site equipment, made available through either telemetry (Jagushte, 2017) and other sensors and scene capture technologies (Arashpour et al., 2020), is on the rise. Data-driven decisions for managing site equipment are highly demanded by contractors' project managers, and in some instances by clients (Niu et al., 2017).

Studies deploying machine learning techniques for various construction management challenges including equipment on site have recently started to emerge (Ghosh et al., 2019). Studies investigating the management of equipment are focussed on understanding usage patterns of equipment, monitoring fuel consumption and carbon emission, managing safety and access to equipment, improving routine and preventive maintenance, servicing strategies (i.e. repair or replacement), procurement strategies (i.e. buy or lease), fleet allocation/deployment, and fleet tracking and localisation (Niu et al, 2017). Current studies explore the use of components such as Building Information Modelling (BIM), Internet of Things (IoT), Radio-Frequency Identification (RFID), and telematics. For clarity, these concepts are defined before exposing the findings of the literature review (Table 1):

- BIM is a digital representation of the physical and functional characteristics of a building over its life cycle (BSi, 2013). BIM is the current expression of digital innovation within the construction sector (Succar and Kassem, 2015).
- IoT is a network of physical objects that contain embedded technology to communicate and sense or interact with their internal states or the external environment (Gartner, 2019). It has a typical four-layers architecture: the perceptual layer (e.g. Wireless Sensor and Actuator Networks – WSAN), RFID, Zigbee, Bluetooth, etc.); the network layer (e.g. communication

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. and Huntingdon, J. (2021). Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction, Volume 124, 2021, 103532, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103532.

networks – satellite network, internet, mobile network, and communication protocols), the support layer (fog computing, cloud computing), and the application layer (IoT applications – smart homes, smart grids, intelligent transportation) (Ali et al., 2016).

- RFID is a wireless technology that exploits the radio frequency (RF) to communicate with uniquely identifiable devices called tags. A RFID system consists of a tag, a reader and a compute device that hosts a database and an application-dependent software package. Each tag has a unique identification number (Fahmy et al., 2019) Connecting RFID reader to the terminal of Internet, readers can identify, track and monitor the objects attached with tags globally, automatically, and in real time, if necessary (Jia et al., 2012).
- Telematics systems refer to any integrated use of wireless communications, vehicle monitoring systems, and location devices to provide real-time spatial and performance data of equipment (Aslan et al., 2012; Said et al. 2014).

The studies reviewed are grouped according to their core applications as follows:

Equipment productivity outputs

Edwards and Holt (2000) estimated the cycle time of excavators, as a measure of productivity, using a 'deterministic' multiple regression model. Data was collected from manufacturers' performance handbooks. Three variables were identified as accurate predictors of cycle time: machine weight, digging depth and machine swing angle. Their model was able to reliably predict cycle times. Schabowicz and Hola (2007) combined queuing theory with artificial neural networks to predict the productivity of a whole system of earthwork machinery made of excavators and haulers with machinery of different sizes and varying hauling distance. A sample of 200 patterns of data were used to train (170) and test (30) the algorithm. Each neural network was generated by computer simulation due to the lack of available real industry data. The results from the simulation confirmed the suitability of the ANN for predicting the productivity of systems of collaborating earthmoving machines. Ok and Sinha (2006) compared linear regression and neural network methods for estimating daily productivity of dozers with the aim of evaluating the potential of using non-linear network analysis for productivity estimation modelling. Neural network was able to produce more accurate results than the regression analysis models. This has led Tam et al. (2002) to develop an ANN, using the same data set of Edwards and Holt (2000), and compare the performance of their ANN with the multiple regression mode of Edwards and Holt (2000). The results showed that the predictive performance of the ANN is superior to the multiple regression models. In the same context, Edwards and Griffiths (2000), driven by the need to improve on the prediction achieved by Edwards and Holt (2000) using multiple regression, have proposed a feed-forward ANN with backpropagation training to predict the hydraulic cycle time of excavators based on data from manufacturers' performance handbooks. They were able to achieve a mean absolute percentage error (MAPE) of 7% (that is, a 14% reduction on the equivalent multiple regression equation).

Although these studies are directly relevant to the aim of this paper, they adopted different metrics for measuring the productivity of equipment such as the cycle time that can be converted into productivity output. Our paper aims to directly measure the volume of earth excavated per unit of time. Existing studies also used different sources for their datasets (e.g. industry data or manufacturers' handbook) while our paper uses data from telematics systems installed on equipment in the field. This also infers a difference in the level of digitalisation and automation of the adopted approach for estimating the productivity. Nevertheless, the positive results achieved in related studies are very promising and encouraging for the aim posed in this paper.

Monitoring safety of workers in proximity of equipment

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. and Huntingdon, J. (2021). Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction, Volume 124, 2021, 103532, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103532.

Kanan et al. (2018) developed an autonomous system for monitoring the safety of workers on site, by combining sensing systems based on three techniques (i.e. radio frequency, directional antennas, and ultrasound waves) with IoT. They installed a sensing system on the rear of site equipment which was used in conjunction with a worker's wearable device that includes a radio transceiver (transmitter/receiver), a wake-up sensor, an alarm actuator, and a General Packet Radio Service (GPRS) module. The combination of the two systems allowed for monitoring, localising, and warning site labourers of proximity dangers. In the same vein, Zhou and Ding (2017) combined RFID, tracking technology, ultrasonic detection technology, and infrared access technology into a three-tier network architecture, to develop a system that assists site workers in changing their risky behaviours and accident avoidance in underground construction sites. The testing of the solution in a metro tunnel construction project showed improved prevention and reduction of accident rates. These two approaches can be exploited to manage safety aspects associated with equipment, but they are not adequate for measuring the productivity of equipment.

Equipment data analytics

Niu et al (2017) proposed a system that enabled the collection of site equipment's operational data and the production of analytics for their management. To achieve this, they proposed the concept of "Smart Connected Objects" where construction resources (e.g., machinery, tools, materials, and even structures) were augmented with sensing, processing, and communication abilities to transform them into smart devices. This edge computing approach was designed to enable equipment to acquire autonomy and awareness so they can interact with their vicinity and surrounding to enable better decision making. The two key components of the proposed systems were: (1) a smart chip that integrates various sensing and communication modules for collecting and exchanging from daily equipment operations; and (2) a data analytics platform for data storage, visualization and analytics. Although machine learning is mentioned as part of the solution, there is little evidence or detail about the techniques implemented.

Lu et al. (2011) used focus groups to explore the application of RFID technologies in the management of material, men and equipment on construction sites. Findings in relation to equipment on site, included potential applications in tracking of equipment and tools, equipment operation permission systems and utilisation records, and equipment maintenance records. However, the authors acknowledged the low adoption of RFID on construction sites and recommended improved dissemination of their benefits and integration with other digital advances (e.g. BIM) as ways for improving diffusion of RFID in site applications.

Equipment health monitoring

Said et al. (2014) developed a telematics-based health-monitoring system to support fleet service managers in decisions about predictive maintenance of equipment. The system consisted of two modules: (1) the health parameters processing and visualization module; and (2) an equipment failure hazard estimation module. At the core of their solution, there are computational algorithms that generate telematics-based fleet use metrics (i.e. maximum coolant temperature, maximum engine oil pressure, maximum engine speed, maximum engine percent torque, etc.). To quantify fleet use, two values are retrieved from each telematics data entry of every piece of equipment: the location of the equipment in terms of the geo zone where it exists; and the time of the reported location. The computational algorithm consisted of an iterative procedural workflow with multiple steps aimed to identify the in-yard and out-of-yard statuses of equipment, infer usage and predict failure hazard. The system tested for both correlations between input variables and failure hazard and accuracy of prediction showed an adequate accuracy especially in later intervals/dates. However, the authors acknowledge the need for further research to improve the accuracy of the predicted failure hazard in earlier dates (i.e. survival time intervals) by integrating additional data from other fleet data sources, such as the engine oil tests from commercial maintenance management software.

Equipment safety monitoring

Zhong et al. (2014) proposed an IoT system called ‘Safety Management System for Tower Crane Groups’. The system is used to detect the operating status of each tower by a set of customized sensors, including horizontal and vertical position sensors for the trolley, angle sensors for the jib and load, and tilt and wind speed sensors for the tower body. Based on the global status data of the whole group of cranes, an anti-collision algorithm was executed to ensure the safety of each tower crane during construction. The algorithm mainly computes the distance from the trolley to the jib and to the coordinates of the tower and applies safe distance rules to issue warning messages. The system enabled the remote supervision of the group of cranes on both personal computers and smartphones. The system was deployed longitudinally on a construction over 12 months by recording the number and types of alarm detected. However, such data was not compared with actual data or data from other sources.

Equipment operational and environmental efficiency

Aslan et al. (2012) developed a system for enhancing the operational productivity of site equipment by combining GPS, WSN, and web applications. Their focus was to provide information on raw data integration and productivity data analysis for the development of fleet management metrics to identify areas of improvement (e.g. guideline to managers, effective equipment operations, resilience and flexibility improvement, and cost reduction. Ahn et al. (2013) developed a system using accelerometers to detect the operational efficiency and infer the environmental performance of site equipment. Their method was based on using vibration signal analysis to detect and monitor the operational status of equipment. Supervised learning algorithms were used to extract various features from the raw accelerometers data and classify them into different equipment activities (working, idling, and engine-off). The system was tested on real construction sites achieving 93% recognition accuracy.

Locating resources on sites

Fang et al. (2016) developed a system by combining BIM and a cloud-enabled Radio Frequency Identification (RFID) system for the localisation of indoor mobile construction resources. BIM was used for system configuration and data visualisation. This visualisation of workers’ locations assists the users of the BIM system in site zoning, security, safety management, and first responder rescue. The system was tested in the field to validate their 3D worker location tracker. Zones over two floors were used to test if the system could identify what zone a worker was in. Where workers took pre-defined paths, the system was able to identify the correct zone with high accuracy (i.e. 88.1%). Limitations of the system included: The RFID network coverage was affected by the range of individual antennas and their layout; system latency was affected by the refresh rate of the RFID readers and the location and signal strength of the router/hotspot; and the RFID was required to be mounted on higher positions such as ceilings or the top of columns to avoid trip hazards. Both the data captured using this approach (i.e. location and position) and the technological setting (i.e. RFID on columns) are not adequate for this paper’s aim that is outdoor and over large construction sites.

Estimating construction costs

Cheng et al. (2010) and Tatari & Kucukver (2011) used neural networks to estimate construction costs for projects. Cheng et al. (2010) use an evolutionary fuzzy hybrid neural network to estimate the cost of projects at the early concept stage. Its variables include a variety of data types, such as manually determined categorical data as well as traditional quantitative data across a range of engineering factors (e.g. temporary construction, geotechnical construction, structural construction, electromechanical construction, etc.). The approach was tested with real industry data in five cases and achieved an overall error in the estimate ranging between 3.3% (case 4) and 20% (case 1).

Tatari & Kucukvar (2011) developed an ANN model to predict the cost premium of LEED certified green buildings based on LEED categories. The input data used for each building consisted of construction year, building type, city, actual construction cost, and scores achieved from LEED categories. The data was collected from online resources for approximately 74 buildings that were used in the ANN development. The standard error of the estimate achieved by the DNN's prediction was 0.61 – the authors acknowledge this as a limitation as a result of the small size of the data set used. Yu and Skibniewski (2010) developed and tested a machine learning cost estimating model, which revealed to be more robust and accurate than traditional elemental cost estimation methods.

Site safety indicators

Poh et al. (2018) tested different machine learning techniques with 13 variables as data entries to develop leading safety indicators for construction sites. An accuracy of 0.78 was achieved in the testing using real industry data. The data used came from one construction company which was acknowledged as a limitation.

Tixier et al. (2016) have both addressed the issue of safety on construction sites using machine learning. They applied two machine learning techniques (i.e. Random Forest, and Stochastic Gradient Tree Boosting) on textual construction injury reports to predict injury type, energy type, and body part and achieved high predictive skills (i.e. Rank Probability Skill Score). In the same vein, Choi et al. (2020) attempted to predict the likelihood of a fatality accident by applying a machine learning technique (i.e. Random Forest) to data from a national database for fatal accidents. They were able to achieve a predictive rate of 91% for classifying workers who may face fatality risks.

Gap in current research

Although this review has excluded studies using scene capturing technologies, a recent extensive review of such technologies can be found in Arashpour et al. (2020). A perusal of this review reveals only three studies focussed on equipment (i.e. Tajeen et al, 2014; Memarzadeh et al., 2013; Azar et al. 2011), all of which focussed on using image or video data sets to detect and recognise equipment on construction sites. One study (i.e. Chen et al., 2020) was focussed on both equipment activity recognition and productivity measurement using a vision-based method. Productivity was measured by estimating the idle time and excavation rate using convolutional neural networks (CNN). The excavation rate (referred to as 'productivity' in the paper) used average bucket payload, total excavation time and total number of cycles, with the first two parameters fixed and the third (i.e. number of cycles) estimated through the CNN model. With the simplified approach (i.e. fixed bucket payload and total excavation time in both ground truth and the estimate), the accuracy of productivity calculation was at 83%.

Based on this analysis and other related reviews such as (Poh et al., 2018), these conclusions can be confirmed: (1) the application of machine learning techniques supported with data from actual construction sites (e.g. through telematics systems) is only emerging for construction management applications generally (as shown in Poh et al., 2018) and for site applications specifically (i.e. focus of this paper); (2) there is a dearth of studies investigating equipment management on construction sites using machine learning and telematics (see Table 1). The specific studies focussing on equipment productivity have used multiple regression analysis. The few available studies proposing machine learning approaches have either: used a manufacturers' handbook or computer simulated data to train and test their models; measured productivity outputs (i.e. cost, volume of earth removed) by predicting the cycle time and maintaining some of the input features unchanged; or focussed on the entire system of equipment on site (i.e. excavators and haulers) which overlook the need to analyse the productivity at the level of each individual equipment.

Table 1. Analysis of related studies

		Technologies			**Techniques**		**Testing / Training approach**	
Paper	**Scope**	**IoT**	**BIM**	**Telematics**	**Machine Learning**	**Other techniques**	**Industry Data**	**Construction site**
Edwards and Holt (2000)	Model and predict productivity and output costs of excavators by estimating the hydraulic cycle time of excavators using multiple regression					✓	✓	
Edwards and Griffiths (2000)	Model and predict productivity and output costs of excavators by estimating the hydraulic cycle time of excavators using ANN				✓		✓	
Schabowicz and Hola (2007)	Model and predict productivity of whole earthmoving machinery systems using queuing theory (excavators and haulers)				✓	✓		
Ok and Sinha (2006)	Compare neural network and multiple regressions technique for estimating of excavator productivity				✓	✓	✓	
Tam et al. (2002)	Predict the productivity of excavators by estimating cycle times				✓			
Fang et al. (2016)	Combining BIM and RFID to produce a tracking system that gives a 3D visualisation of worker location		✓			✓	✓	✓

Cheng et al. (2010)	Use neural networks to improve conceptual cost estimate precision				✓			
Yu and Skibniewski (2010)	Develop neurofuzzy system to estimate the cost of residential building				✓			✓
Kanan et al. (2018)	Create a wearable device and associated IoT systems that will alert a worker to hazardous areas	✓					✓	
Niu et al. (2017)	Integrate IoT to site equipment to allow data collection, management, and analysis	✓			✓		✓	✓
Said et al. (2014)	Employ telematics data to assist decision-making involving fleet use assessment and equipment health monitoring			✓			✓	
Tatari and Kucukvar (2011)	Use neural networks to predict the cost premium of LEED certified green buildings				✓	✓		
Zhong et al. (2014)	Use IoT to collect status data of crane arms and build an anti-collision algorithm to ensure the safety of each tower crane during construction	✓				✓	✓	✓
Zhou and Ding (2017)	Use IoT and RFID to generate a warning system to alert workers to potential hazards	✓				✓	✓	
Ahn et al. (2013)	Use accelerometers to analyse operating efficiency for monitoring environmental performance of equipment	✓			✓			✓
Poh et al. (2018)	Develop leading indicators to classify construction sites according to their safety risk				✓			
Cheng et al. (2020)	Use text mining to classify construction site accidents				✓		✓	
Tixier et al. (2016)	Use text mining to predict injury by predicting predict injury type, energy type, and body part				✓		✓	
Choi et al., (2020)	Develop a prediction model to identify the potential risk of fatality accidents at construction sites				✓		✓	

3 Interviews of experts

To understand the challenges affecting the operation of equipment on site, 13 experts from the Costain Skanska Joint Venture (CSJV) at the High Speed Rail 2nd Phase (HS2), one of the largest infrastructure projects in Europe, were interviewed. A snowball and a purposeful sampling approach were used to identify the 13 participants. The participants were all in relevant roles including project director, project manager, plant and equipment manager, organisational development professionals (lean practitioners), site foreman, surveying and monitoring staff, health and safety professionals and compliance managers. They were all directly involved or affected by the operation of equipment on site.

As the problem is multifaceted and the research team had prior knowledge from securing industrial research funding to investigate the topic, an open unstructured interviewing process was adopted. This process is also referred to as in-depth interviewing. It allows a deeper understanding of a topic compared to other interviewing approaches even in cases when the researcher has a prior understanding of the topic. Interviewees can respond freely about events, behaviour and beliefs related to the questions (Saunders et al., 2015). The researchers can be guided by a list of initial of open questions while preserving the opportunity for new questions to emerge and for existing questions to be changed, reframed or eliminated. One of the risks with open unstructured interviewing process is that it can drift from its core focus and hence, it generally requires more experienced researchers. In this research, the interviewing team consisted of three individuals including a research fellow, an assistant professor and a full professor, and they were also supported by their industry collaborators who are collaborating partners in the funded research project.

The two main topics used by the researchers as a point of departure were:

- Understanding current processes for managing equipment on site and the key challenges faced; and
- Understanding current processes for monitoring and controlling the performance of the excavation programme of work and the challenges faced.

The discussion that unfolded under each question led to further sub-themes and uncovered specific challenges. For example, the first question quickly moved into the challenges of understanding equipment usage and their productivity which is the main challenge for earthwork contractors and subcontractors. As this was the core topic of the study, many follow-up questions (e.g. how productivity is currently measured; how accurate the current measurement is; what technologies are used; what data is captured; who owns the data and who requires such data) were asked by the researchers. Findings covering these items are presented in the next sub-section. Similarly, the second question has uncovered several sub-themes which are not covered in this paper as they are outside of the scope.

In addition to identifying ‘sub-themes’, the discussion around the two questions has also identified ‘new themes’ representing management functions related to equipment on site. These are only summarised hereafter for completeness, but their details are not in the scope of the paper: (1) Management of statutory requirements (i.e. monitoring and reporting undertakings and assurances) which includes the identification and notification of breaches (e.g. noise level from machinery, access to properties is maintained and no infringement access to certain properties, equipment emissions) and requires a self-management process with auditable log; (2) Traffic management planning and offsite logistics: A construction programme is bound by traffic management controls, e.g. transport to and from site within set hours, not permitted waiting outside of site. Logistics need to be coordinated in a timely manner with programme task readiness and operator availability; and (3) Health and safety and compliance.

The experts thought the current process is not very accurate as the currently used remote sensing approach (Light Detection and Ranging – LIDAR) do not consider backfilling and measurements have to rely on manual estimates by site supervisors. They argued that the threshold for usefulness of prediction should be around 80%, tolerating a 20% error.

The findings about the main challenge (i.e. see Section 3.1) were validated through data saturation. In this instance, saturation is achieved when the themes identified fully fit the emerging concepts and categories from new interviews hence, new information or themes are observed in the data (Guest et al., 2006). As data saturation addresses both the quantity and the quality of information, it has also assured the accuracy of the previously captured data.

3.1 Measuring equipment productivity: current approach and challenges

The key challenge addressed in this paper is the measurement of equipment productivity and usage. The interviews helped understand the current process used within the industry and the challenges faced. These findings are summarised hereafter.

Contractors require accurate reports of excavated soil quantities to clients as these quantities are used for compensating contractors and subcontractors under certain procurement arrangements such as the ‘cost-plus’, meaning that the contractor is charged for the volume of work completed as the project progresses. The risk balance in a cost-plus is skewed in favour of the contractor. This procurement arrangement is selected when they are generally many unknowns in earthwork projects. For example, the sensitivity and complexity of certain project sites (central London) mean that the bill of quantities are likely to be subject to underestimation. However, to address this unbalanced distribution of the risk, clients mandate that the contractor should be able to accurately report volumes of earth excavated and provide evidence of how efficiencies are being driven.

Respondents argued that a proxy measure of productivity in such circumstances is the actual volume of earth removed. Moreover, the actual volume, according to the experts, can be used in programme control when comparing them against estimated (budgeted) quantities used in the bill of quantities to identify areas affected by productivity issues.

The scope affected by this measure extends to the entire programme of demolition and excavation work. The key individuals involved are the equipment operator, the equipment hire company (i.e. provision of telematics data about equipment operation), the site agent (i.e. collection of data for reporting utilisation and productivity of equipment), the site superintendent (i.e. acceptance of equipment and labour for fitness to operate on site), the project manager (i.e. decision-making on equipment and labour), and the waste removal company involved in environmental compliance (i.e. transport and disposal).

In current practice, contractors have attempted to use existing telemetry systems that are fitted to plants and equipment to address this need. However, the volume of earth excavated is not directly captured by or easily inferred from telematics systems’ data, additionally there could also be issues of inconsistency and incompatibility in the data set captured by telematics. These are caused by different temporal reporting frequencies by the various original equipment manufacturers (OEM), different measurement approaches and units (e.g. defining idle times differently), different timing for transmitting the data, and limited availability of telemetry systems on equipment of a lower weight. To avoid these limitations, construction sites use alternative approaches to estimate the volume of soil removed. They employ supplementary measurement devices including laser scanning by drones to measure topology changes from which volumes of soil removed can be estimated. Manual measurements from drones are reported daily for each work zone requiring labourers on site to estimate volumes excavated. However, the use of drones only can provide high granularity measurement which means that productivity and utilisation of individual equipment is not possible to detect. This capability is important in large infrastructure projects seeking to demonstrate efficiencies.

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. and Huntingdon, J. (2021). Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction, Volume 124, 2021, 103532, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103532.

The inability of accurately estimating volume of soil excavated was also considered by the experts as a barrier to effective control of existing project and to accurate costing and budgeting of future projects and project stages. Indeed, monitoring of programme performance is carried out at daily and weekly intervals by drawing data from disparate sources to populate a ‘control board report’; these data sources included: site diaries, LIDAR drone surveys, waste-transfer notes, and baseline and actual programmes.

This research explores an alternative approach to this use case. It proposes the application of DNN to estimate the productivity of equipment as a function of the volume of earth removed. The development and testing of the DNN method is presented in the next section.

4 Development and optimisation of the DNN model

An approach using machine learning to automate the measurement of volume of earth excavated is proposed in this section. The approach implements DNN and uses data from the telematics systems that are available on excavators. If successful, the proposed approach would change the current approach which is onerous and require the use of a combination of data sources from drone LIDAR scanning, load counts, and vehicle weighting. The next subsections describe and discuss the steps involved in developing and testing the proposed DNN model.

4.1 Feature engineering: vehicle telemetry data

Telematics systems provide data that can be used as training, validation and test data sets when building the DNN. Equipment telemetry data were made available from three hiring companies (via their respective OEM telemetry services) who were contract hirers for the current case study site.

DNN was selected for a number of reasons: the challenge investigated is a supervised learning problem with numeric predictions, so regression-based learning within the model chosen was a priority; the data available as evidenced later is noisy and a machine learning approach such as the DNN is more adequate compared to other statistical methods such as linear regression; and the data used had many variables, so ensuring that the network was a multi-layered perceptron with a complex input space is important.

The DNN model was trained using all relevant available data. A complete list of the available data types can be found in Appendix A. However, as the DNN model failed to converge, the data was refined. Some data was removed for being non-numeric, or incomplete, whilst other parameters were removed through subset selection, whittling down the number of variables until the model converged effectively (see Appendix A for details for a summary of the data removed and selected). Table 2 gives a feature analysis of the retained data. This approach is in line with other studies such as (Schabowicz and Hola, 2007). However, (Schabowicz and Hola, 2007) predate the emergence of telematics systems and many of the identified factors included in their method (i.e. hauler loading time, hauler work cycle time, excavator bucket capacity, hauler loading platform capacity, hauler driving speed, excavator bucket working speed, the kind of road surface, the category of the soil, the number of excavators working in the system, the number of haulers working in the system, the excavator work cost and the hauler work cost) are not readily detectable by telematics systems.

Table 2. Feature analysis of telemetry data (test 3a)

Feature	Unit	Count	Mean	St dev	min	*25%	*50%	*75%
Volume excavated	m3/d	1184	16.50	27.16	0.00	0	1.25	47.55
Fuel rate	litres/hr	790	10.64	8.98	0.00	6.30	8.27	10.60

Vehicle weight	tonne	1184	20.18	10.57	13.00	13.00	21.00	22.00
Bucket volume	m3	1184	0.75	0.44	0.43	0.43	0.86	0.86
Total fuel consumed	litres	790	60.70	65.42	0.00	23.00	48.50	69.00
Engine on time	Hrs	790	4.89	3.06	0.00	1.73	5.93	7.33
Travel time	Hrs	790	0.50	0.48	0.00	0.15	0.32	0.73
Engine on (no dig)	Hrs	790	4.88	3.04	0.00	1.84	5.88	7.30
Engine on (no move)	Hrs	790	2.61	1.98	0.00	0.87	2.53	3.88
Digging	Hrs	790	2.55	1.77	0.00	0.95	2.62	3.88
Slew arm swing	Hrs	790	1.72	1.35	0.00	0.51	1.59	2.73
Not operating	Hrs	790	2.66	1.73	0.00	1.21	2.72	3.93

*includes all zero target values (volume excavated), these were not included in the processed training data.

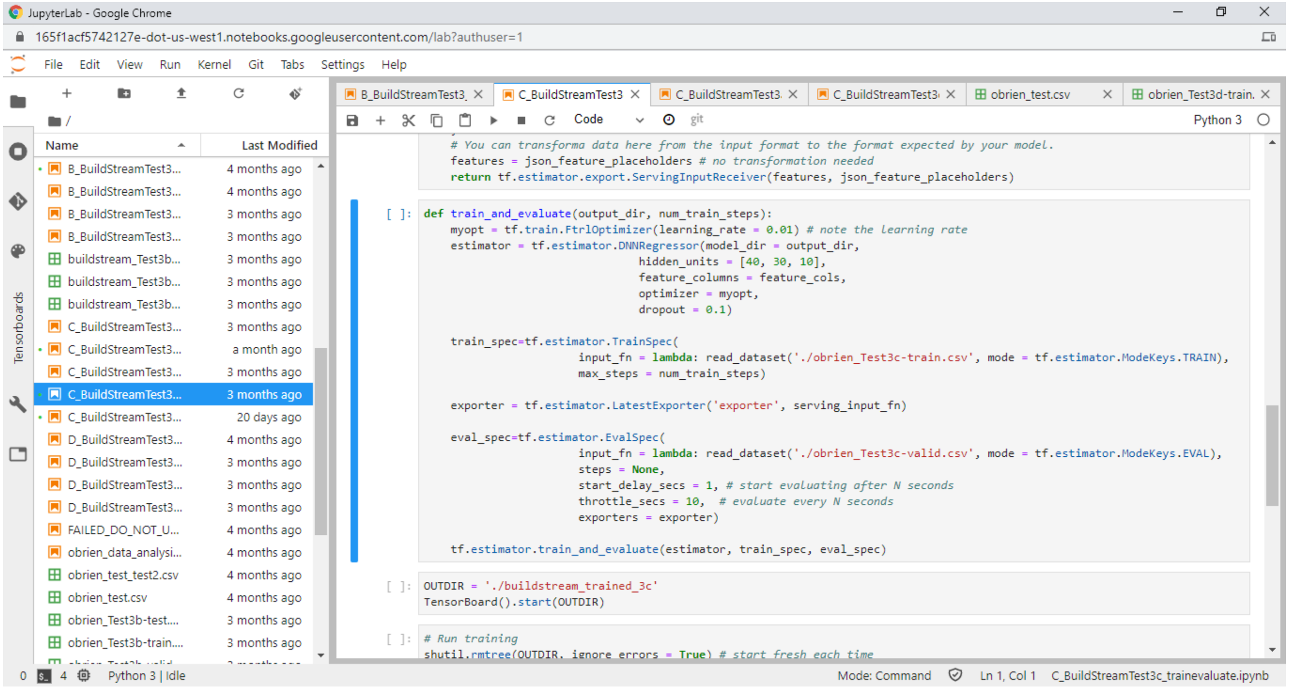
Table 3 presents a density distribution for the training, validation, and test data. The minimum threshold for data volume and variety for training a model was a time series that captured seasonal trends (i.e. construction equipment performance is dependent on ground conditions as a result of weather), a variety of equipment types (based on the vehicle weight: those represented in the data set were 13, 21, 22 and 49 tonnes), and sufficient samples representing each tenth percentile of volume excavated (i.e. 10th percentile, through to the 90th percentile); a minimum of five input samples was required at each percentile interval. For the purpose of feasibility testing, data instances were created as static files and features were presented to the ML model as CSV instances (Figure 1). However, in future this offline approach can be replaced with SQL data storage and run data processing scripts 'Dataproc' to present the data for ongoing AI model training.

The DNN model was developed within a cloud-based data pipeline. The Google Cloud Platform (GCP) was selected as it offered a way of ingesting, preparing, and storing data for machine learning functionality. The DNN was built within the Jupyter Python environment using TensorFlow. Its 'Estimator' function allowed the focus to be on building models quickly, which then enabled rapid checkpointing of models and evaluation of their performance. Prediction performance was improved through hyperparameter tuning by modifying the network architecture, selecting different DNN regressors and different activation functions (i.e. regularization - ReLU). This process is demonstrated in the next subsection.

Table 3. DNN feature input data density distribution

Percentile	**Training set**		**Validation set**		**Testing set**	
	Volume (m^3/day)	**Samples (nr)**	**Volume (m^3/day)**	**Samples (nr)**	**Volume (m^3/day)**	**Samples (nr)**
10	0.13	56	0.16	13	0.15	11
20	0.69	56	0.56	12	0.66	10
30	3.07	55	2.27	12	4.45	10
40	8.62	56	10.57	12	12.26	10
50	18.93	55	15.83	12	21.60	10
60	42.56	56	28.78	13	43.56	10
70	84.95	56	74.16	12	74.82	10
80	150.72	55	155.91	12	149.68	10
90	252.53	57	241.07	12	234.82	10

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. and Huntingdon, J. (2021). Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction, Volume 124, 2021, 103532, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103532.



Name	Last Modified
B_BuildStreamTest3...	4 months ago
B_BuildStreamTest3...	4 months ago
B_BuildStreamTest3...	3 months ago
B_BuildStreamTest3...	3 months ago
buildstream_Test3b...	3 months ago
buildstream_Test3b...	3 months ago
buildstream_Test3b...	3 months ago
C_BuildStreamTest3...	3 months ago
C_BuildStreamTest3...	a month ago
C_BuildStreamTest3...	3 months ago
C_BuildStreamTest3...	3 months ago
C_BuildStreamTest3...	20 days ago
D_BuildStreamTest3...	4 months ago
D_BuildStreamTest3...	3 months ago
D_BuildStreamTest3...	3 months ago
D_BuildStreamTest3...	3 months ago
FAILED_DO_NOT_U...	4 months ago
obrien_data_analysi...	4 months ago
obrien_test_test2.csv	4 months ago
obrien_test.csv	4 months ago
obrien_Test3b-test....	3 months ago
obrien_Test3b-train....	3 months ago

```
    # You can transforma data here from the input format to the format expected by your model.
    features = json_feature_placeholders # no transformation needed
    return tf.estimator.export.ServingInputReceiver(features, json_feature_placeholders)
```

```
[ ]: def train_and_evaluate(output_dir, num_train_steps):
         myopt = tf.train.FtrlOptimizer(learning_rate = 0.01) # note the learning rate
         estimator = tf.estimator.DNNRegressor(model_dir = output_dir,
                                 hidden_units = [40, 30, 10],
                                 feature_columns = feature_cols,
                                 optimizer = myopt,
                                 dropout = 0.1)

         train_spec=tf.estimator.TrainSpec(
                        input_fn = lambda: read_dataset('./obrien_Test3c-train.csv', mode = tf.estimator.ModeKeys.TRAIN),
                        max_steps = num_train_steps)

         exporter = tf.estimator.LatestExporter('exporter', serving_input_fn)

         eval_spec=tf.estimator.EvalSpec(
                        input_fn = lambda: read_dataset('./obrien_Test3c-valid.csv', mode = tf.estimator.ModeKeys.EVAL),
                        steps = None,
                        start_delay_secs = 1, # start evaluating after N seconds
                        throttle_secs = 10,  # evaluate every N seconds
                        exporters = exporter)

         tf.estimator.train_and_evaluate(estimator, train_spec, eval_spec)
```

```
[ ]: OUTDIR = './buildstream_trained_3c'
     TensorBoard().start(OUTDIR)
```

```
[ ]: # Run training
     shutil.rmtree(OUTDIR, ignore_errors = True) # start fresh each time
```

Figure 1. Equipment telemetry presented as training inputs to machine learning on Google Cloud using Python (Jupyter Notebooks)

4.2 Optimisation algorithms

One of the areas of deep learning that this research has investigated is the evaluation of the effects of three different gradient descent (GD) optimisation algorithms to improve the prediction accuracy. Gradient descent (GD) is the method by which the error is minimised and fed back into the neural network to update weights accordingly. Three different gradient descent optimisation algorithms were investigated in the development of the DNN for prediction of the volume of earth removed. These optimisation algorithms can be used to improve the prediction capacity of the trained neural network.

In its simplest form, gradient descent uses a fixed learning rate when 'descending' the error curve. However, the ability of the neural network to seek out the absolute minimum depends on the non-linearity of the input features being presented to the network versus the desired outputs. The descent curve is prone to 'early stopping' of the training, where the apparent minimum error is a local minimum and not the global minimum. To overcome entrapment in local minima, adaptation of the initial learning rates can be applied. Three different gradient descent algorithms were tested: Adam, AdaGrad, and Follow the Regularized Leader (FtRL). This paper is not intended to provide an exhaustive series of experiments of optimization; however, it is important to understand how sensitive the trained DNN is to learning rate. The differences between the three optimisers are as follows:

- Adam optimisation is a form of stochastic gradient descent (a variation of gradient descent) that iteratively updates network weights by using an exponential moving average of the gradient and the squared of the error gradient (Brownlee, 2017a).

- Adagrad is an algorithm which adapts the learning rate at different rates depending on the most frequent and infrequent features within the input data, where frequent input features are updated with smaller learning rates whereas infrequent features are updated with larger learning rates (Ruder, 2016).
- FtRL is a classification algorithm that combines L1 (Lasso Regression) and L2 (Ridge Regression) regularisation and adaptive learning rates. L1 and L2 are algorithms where different penalties are applied to the loss function depending on whether losses are small (close to zero) or very large. The principal is that in order to avoid overfitting, the penalty function shrinks less important features' coefficients to zero, removing their influence within training altogether. It effectively automates the selection of input features by recognising redundancy in the input space. The difference between L1 and L2 is that L1 applies a penalty term using the 'absolute value of magnitude', whereas L2 adds the 'squared magnitude', to the penalty terms (Nagpal, 2017).

Mini-batch gradient descent

Gradient descent (GD) has different variants in terms of the number of training samples used to update the model. One of the GD-based techniques is batch GD where the loss function is calculated after each pass of the entire training data set. In large data sets, batch GD can lead to slow performance and poor fitting. Mini-batch GD is another variation of GD, which can outperform batch gradient when a model uses large datasets. Mini-batch GD divides the training data set into small set of samples, called "mini-batch" (Brownlee, 2017b). The batch size is a hyperparameter of gradient descent that controls the number of training samples to work through before the model's internal parameters are updated. The current research adopted mini-batch GD and tested the sensitivity of the DNN performance using different batch sizes.

A series of gradient descent tests were performed to evaluate the most appropriate algorithm, the starting learning rate, and the mini-batch size for the equipment telemetry data for predicting excavation volumes per day. Each GD algorithm (Adam, Adagrad and FtRL) was applied to the same set of training data. Each run was repeated for a starting learning rate of 0.001, 0.01 and 0.1, respectively. The root-mean-square error (RMSE) was found for each combination of parameters as displayed in Figure 2. The smallest RMSE was 5.015, indicating that the best-performing model was an Adam optimiser with a batch size of 70, and a learning rate of 0.01. This was considered adequate, given the limited training data set (700 samples), the type of application (construction earthwork), presence of manual work (i.e. archaeologists) alongside the equipment on the selected case study site. Nevertheless, larger data samples (~1000 data sets) from different sites will be needed in future, especially when prediction will be made based on either real time or batch processing.

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. and Huntingdon, J. (2021). Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction, Volume 124, 2021, 103532, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103532.

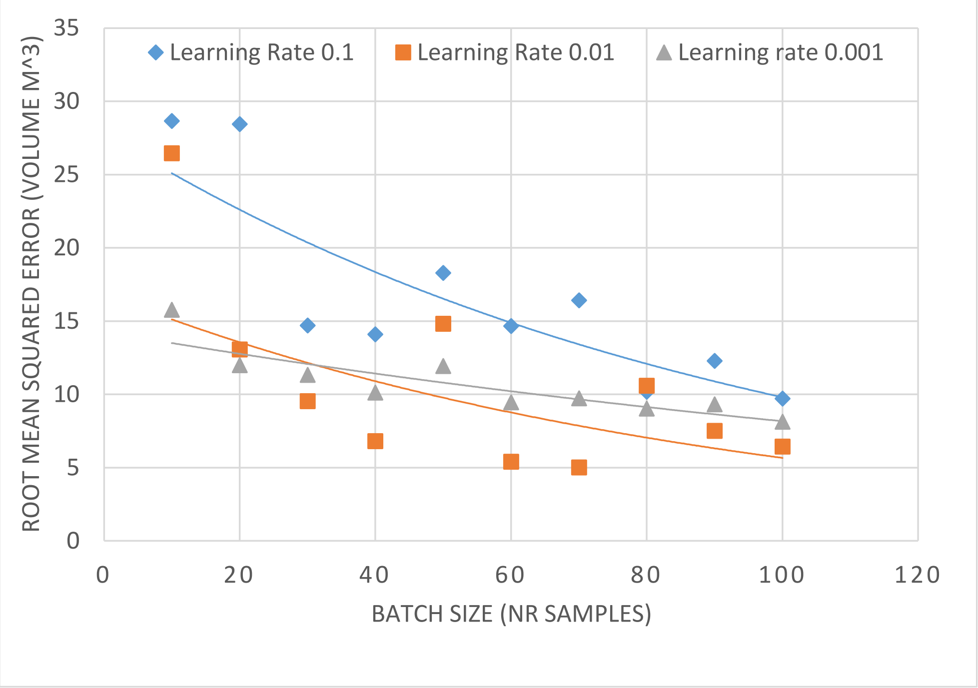


Figure 2. Gradient descent optimization using 'Adam' optimizer

Following the identification of the preferred gradient descent algorithm and batch size, the next step is to determine the optimal DNN architecture. This includes finding the optimal number of hidden layers and number of nodes. A series of tests were run using different configurations of layers as listed in Table 4. The process starts with a high number of connections and progressively reduces complexity by decreasing the number of hidden nodes to avoid overfitting problems until identifying the architecture that increases the generalisation capability without decreasing the accuracy of the model. The accuracy performance was measured in terms of root mean square error (RMSE). The results show that the best performing DNN architecture was two hidden layers with 50 and 25 nodes in the first and second hidden layers, respectively. The result of the first round of gradient descent tests was improved upon marginally by achieving a RMSE of 4.410 (5.015 before architecture tuning). The resulting DNN model is depicted in Figure 3.

Table 4: The RMSE results of architecture parameters on the machine learning model. Note: All architecture configurations run using 'Adam' optimizer, batch size 70, and starting learning rate 0.01

3 hidden layers	**RMSE**	**2 hidden layers**	**RMSE**
15, 7, 3	11.035	--	--
30, 15, 5	6.190	30, 15	6.799
40,30,10	12.856	40, 20	5.847
50, 30, 20	9.318	**50, 25**	**4.410**
60, 30, 10	6.304	60, 30	7.126
60, 40, 20	9.060	--	--

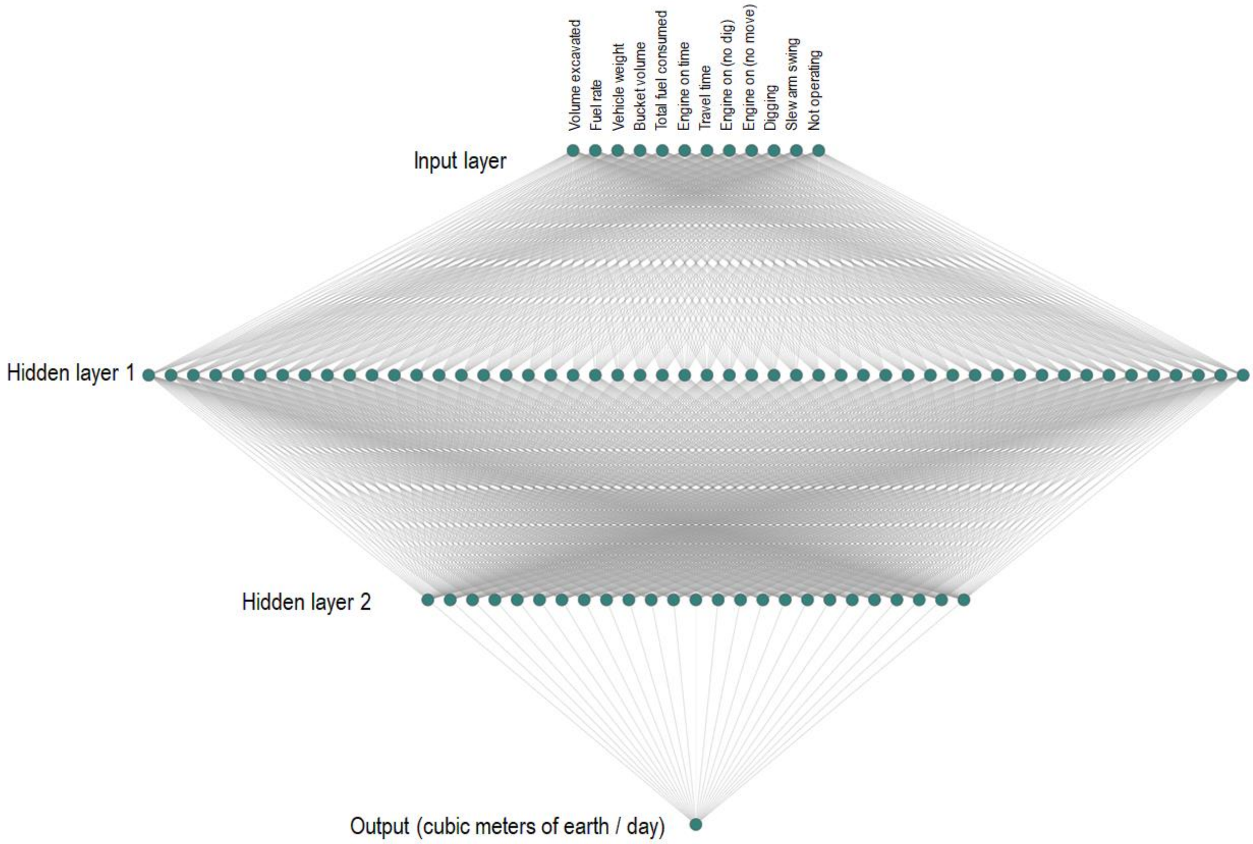


Figure 3. DNN model for predicting volume of earth excavated

5 Testing and evaluation of the DNN model

To appraise the performance of the DNN model, four different metrics were considered (Table 5).

The mean absolute error (MAE) is a loss function used for regression. It takes all the real data and their associated predictions and finds the absolute error for each pairing. A disadvantage of MAE is that the gradient magnitude is not dependent on the error size which means that in some cases the gradient magnitude could be larger even when the error is smaller. Hence, it is difficult to get a sense of scale from the results produced (e.g. is 10 good or bad?). The MAE is usually only useful when used to compare two different predictive methods: the one with the smaller MAE is the one that is closer to the ground truth. In conclusion, MAE values are difficult to interpret in isolation and are only practical as comparative measures.

The mean absolute percentage error (MAPE) is one is one of the most popular measures of the forecast accuracy. MAPE is a measure that seeks to fix the issue of isolated values seen in the MAE by presenting the error as a percentage. Percentages are a preferable way of representing the error in this context and remain comprehensible without another value to compare against. MAPE has some key disadvantages: it produces infinite or undefined values when the actual values are zero or close to zero although higher values generally indicate less useful models; it favours negative errors and cases where the forecasts are lower than the recorded values. As some of these two circumstances (i.e. actual value close to zero, and low forecasts) are very likely in earthwork activities, this method was not selected for use in this research.

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. and Huntingdon, J. (2021). Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction, Volume 124, 2021, 103532, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103532.

The weighted absolute percentage error (WAPE) is a method that seeks to rectify some of the problems identified with the MAPE. Although it can also produce value between zero and infinity, the total difference is divided by the total actual values before converting it into a percentage which removes the issue of having to divide by zero (at least, it does in a data set in which all values fall in the $\mathbb{R}^+$ set). For these reasons, the WAPE was considered as the most appropriate measure of model accuracy for this research. In addition to the WAPE, the study also adopts the coefficient of determination (R^2), a widely used method to assess the quality of predictions by measuring the overall goodness of fit of a model.

Table 5. methods used to evaluate performance of predictions

Method	**Formula**	
Mean absolute error (MAE)	$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert v_i - f_i\rvert$	N = The number of samples. v_i = measured volume of sample i, in m^3. f_i = predicted volume forecast associated with sample i, in m^3. V = the mean of all measured volumes v_i, in m^3.
Mean absolute percentage error (MAPE)	$MAPE = 100 \times \frac{1}{N}\sum_{i=1}^{N}\left\lvert\frac{v_i - f_i}{v_i}\right\rvert$	
Weighted absolute percentage error (WAPE)	$WAPE = 100 \times \frac{\sum_{i=1}^{N}\lvert v_i - f_i\rvert}{\sum_{i=1}^{N} v_i}$	
coefficient of determination (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{N}(v_i - f_i)^2}{\sum_{i=1}^{N}(v_i - V)^2}$	

Data over a 21-day period was used to test the DNN prediction of volume of earth removed. The results of this are displayed graphically in Figure 4.

The R^2 achieved by this study is 0.87 which is very adequate for the problem addressed in this study as evidenced later on in this section. The WAPE, evaluated on data points from 21 days, is: 30.36%. This is 10.36% below the threshold of precision (that is 20%) expected by industry expert. The precision level achieved can be judged using two approaches: (a) inspecting the specifics of the prediction over the 21 days, and (b) comparing the performance achieved in this study with those of other studies addressing the same question. However, the literature review evidenced the lack of studies measuring the same output and using similar feature inputs and experimental settings. Hence, a comparison will be made with some studies addressing 'similar' construction applications to insinuate some general conclusions about the precision of the prediction achieved.

The first approach reveals interesting findings in relation to the precision level. When the prediction results are analysed over 21 days (Figure 4), a clear difference can be seen between days with high excavation volumes and days with low excavation volumes. Accurate and better predictions were consistently obtained in days with high excavation volumes. Days when excavation volumes are low indicate likely involvement of manual work (i.e. archaeologists) alongside some of the equipment which causes gaps in telematics data that adversely affect the accuracy of predictions. This means that the findings of prediction are adequate, but they are more accurate in a highly mechanised approach (i.e. earthwork tasks with equipment mainly and limited human task interventions) compared to a mechanised-manual mixed working environment.

To judge the precision of predictions according to the second approach, there is a need to identify studies that predicted the same output using the same or other machine learning techniques. However, as demonstrated in the literature review the existing studies: did not measure directly the volume of earth excavated but have opted to predict different productivity measures such as the hydraulic cycle of excavators; used data sets from manufacturers' handbooks or computer simulated data which is not affected by noise and gap issues, or adopted a totally different approach such as a vision-based method with fixed input features such as bucket payload, etc. These are significant differences which infer that the mere comparison of outcomes without considering their surrounding implementation environment should not be the only approach to invoke ultimate conclusions about the performance achieved in this study. Moreover, Several error metrics used to evaluate the prediction accuracy such as RMSE, Mean-Square Error (MSE) and correlation coefficients (R), and there is not a consensus on a common error metric which make cross-reference difficult to conduct (Tian et al., 2020). Nevertheless, despite these general challenges and the peculiarities of this study's prediction, the precision levels attained in this paper were found to be comparable to those in 'similar field'. For example, this study's R^2 value (0.87) is very close to the R^2 (that is 0.92) achieved in a study using DNN to predict the penetration rate of tunnel boring machines. An interesting comparison is with the work done by Rashid et al. (2019) who had to implement data augmentation techniques to generate a synthetic training data set to train the model with a large data volume for better performance. Before data augmentation, the accuracy of their model in predicting the activity recognition of excavators was 63.3%. Based on these comparative data, the accuracy achieved in this study is acceptable, but it can be improved as explained in Section 7.

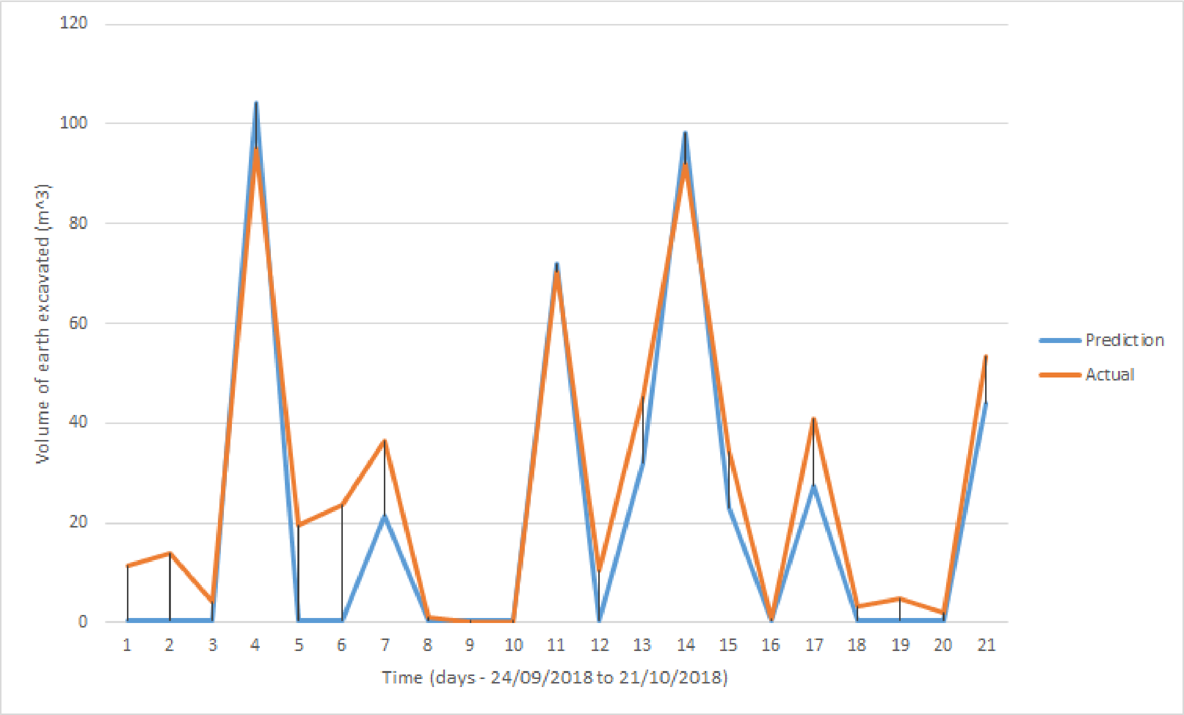


Figure 4 DNN prediction of excavated volume for a 21 day multiple equipment digging period

6 Developing benchmarks and benchmarking for earthwork projects

Using predictions to develop benchmarks and perform benchmarking is of paramount importance within the construction sector generally and in infrastructure projects particularly. Benchmarking is an approach that enables the comparison of performances of processes, activities and deliverables within and between projects and organisations over time (Kassem et al., 2019). It is used in industrial, commercial, and technological environments to compare measurements of performance and identify areas of improvement and good practice. In the context of this research, predictions made and recorded over time can form the basis for producing benchmarks which become points of reference against which equipment performance measurements can be compared. These could be done at different levels including the whole earthwork site, work zone or individual equipment, and can be used to compare performance within the same project or across projects. During interviews with industry experts, the benchmark required in earthwork projects were identified alongside the desired frequency for carrying out the benchmarking (Table 6).

Table 6: Performance measures used by industry experts when evaluating the progress of excavation.

Benchmark	**Units**	**Frequency of benchmarking**
Total volume excavated	m^3	Monthly
Total volume filled in	m^3	Monthly
Cost of excavation	$£/m^3$	Monthly
Load counts	Integer counts	Daily

The ANN predictions performed and demonstrated earlier are for the total volume excavated as required by the industry experts in Table 6. However, as the total volume excavated is an absolute measure that is affected by individual project characteristics (e.g. size, type of soil, etc.), it would be appropriate for use within large and individual earthwork projects (intra-project benchmarking) but not across projects (inter-project benchmarking). A relative measure that allows for easier comparison across projects is desirable. Indeed, according to the Infrastructure and Projects Authority (IPA, 2019), in their Best Practice in Benchmarking report, the development and application of benchmarking should facilitate the consistent collection, collation and sharing of comparable data across infrastructure delivery organisations may be a preferable benchmark to carry forward. In line with this principle, this research proposes a bottom-up approach to benchmarking which adopts the excavation rate as the performance measure. The excavation rate is the volume of soil excavated per unit of time (e.g. hour), and being a bottom-up measure it can be also calculated for each individual equipment type. The benchmark equation for the excavation rate of a set of machines, B, is given by this equation:

$$B = \frac{\sum_{i=1}^{N} \left(\frac{v_i}{d_i}\right)}{N}$$

with the following variables:

- N = The number of samples.
- v_i = The volume of sample i, in m^3.
- d_i = The digging hours for sample i.

The results of this benchmark measure over the training data – both over all the vehicles and each cluster by vehicle weights can be found in Table 7. The distribution of the training data sample across the different equipment clusters is included to show that each vehicle weight has a reasonably high number of samples, directly obtained from telematics system.

The data show that the excavation rate varies between vehicle weights and is not proportional to the equipment weight. Hence, the data is also counter intuitive as some of the smaller excavators have a higher excavation rates than some of the larger ones in some instances. This can be attributed to either (1) the nature of the testing site being an archaeological site requiring a mixed mechanised-manual work in which smaller excavators are more adequate than the large ones; (2) issues in the productivity of some of the larger excavators caused by breakdown, idling time and other site (e.g. soil type) or weather conditions. While these findings expose further challenges to the attainment of more accurate prediction, they also re-confirm that the proposed approach is more adequate for a highly mechanised environment (i.e. excavation work with equipment predominantly and limited human interventions) compared to a mixed mechanised-manual working environment.

Table 7: The rate of excavation for the recorded data, in cubic meter per hour, for each vehicle weight, and the number of samples for each weight.

Vehicle weight (tonnes)	$B\ (m^3/h)$	**N**
All	22.034	790
13	18.873	376
21	15.875	152
22	7.166	189
49	89.634	73

Using the proposed benchmark measure to build performance datasets over time can help to understand how well a zone, site, vehicle or set of vehicles is performing. Further implications are discussed in the next section.

7 Discussion and future developments

Inefficiencies in the management of equipment fleets in earthwork activities on infrastructure projects can have significant consequences on economic and technical feasibility of the projects and the financial health of contractors and subcontractors. The ability to measure and benchmark their performance is key to improving their management. One important performance measure identified in this study is the ability to estimate the volume of earth excavated. Not only was this identified as a research gap in the existing literature, but also discussion of current practice for managing equipment with industry experts revealed some important challenges such as manual and time-consuming workflows and inaccurate measurements.

The review of existing literature demonstrated that a limited number of studies have focused on equipment productivity; a dearth of studies have attempted to measure the volume of earth excavated; and only a few studies have used a combination of machine learning and telematics. To address this gap, the aim of this paper was to develop and test a DNN model to estimate the productivity of excavators which can also be used to develop a benchmarking approach. One of the main performance measures used to estimate the productivity of earthwork is the volume of earth removed or shifted in a unit of time (e.g. hour, day or week).

Each step of the DNN development and testing was exposed and discussed including feature engineering; development and optimisation of the DNN model, and evaluation of its performance and its use for benchmarking purpose. This section discusses the key findings and challenges from each step including future improvement or alternative developments.

The process of feature engineering used telemetry data that included 700 samples (final selection) covering three weeks of site activities. There were two major challenges in working with this data.

Firstly, there were inconsistencies in data structure between telematics systems from different OEM providers and inconsistencies in the time intervals between data uploads across the different providers. This meant that the disparate datasets could not be merged effectively, so only one data OEM provider was chosen for development and the work has been done for the equipment provided by this provider. The final data set was chosen due to the data having the highest temporal recording (daily) and the widest array of data fields to test for important feature inputs. These findings provide practical evidence for the need for streamlined IoT devices in order to create the necessary data pipelines that can be managed by cloud-computing as set out by the Association of Equipment Management Professionals 2.0 Standard (Rehman, 2019). These requirements have recently evolved into an ISO standard (i.e. ISO/TS 15143-3:2020) that specifies a common data format to standardise the retrieval of telematics data from mixed equipment. Twenty common parameters are part of the standard including asset identification, location, operating hours or miles, fuel burn, engine temperatures, fuel level, idle time and average power percentage (ISO, 2020).

The development of the DNN started with the testing of different gradient descent optimisation algorithms (i.e. Adam, AdaGrad, and Follow the Regularized Leader) using different learning rates and batch sizes. The configuration of the DDN (i.e. Adam optimiser with a batch size of 70, and a learning rate of 0.01) that achieved the lowest RMSE was selected. The smallest RMSE achieved was 5.015 which was considered adequate given the limited training data (700 data sets), the type of application (construction earthwork), presence of manual work (i.e. archaeologists) alongside the equipment on the selected case study site. Despite the accepted levels of prediction performance achieved, larger data samples from different sites will be needed in future, especially if prediction has to be made based on either real time or batch processing. The optimisation also aimed to identify the optimal network architecture. The best performing DNN architecture was with two hidden layers with 50 and 25 nodes in the first and second hidden layers, respectively. This agrees with the literature on ANN where networks with two hidden layers are often found to be better generalisers although the actual degree of improvement is case dependent (Thomas et al., 2017).

The R^2 and WAPE, selected as a method to measure the accuracy of the DNN predictions, were estimated at 0.87 and 69.64%, respectively, from data points over 21 days. While the R^2 is comparable with the prediction performance attained in studies addressing similar construction challenges, the WAPE is 10.36% below the accuracy threshold (i.e. 80%) required by industry experts. However, when considering the special characteristics (e.g. training data from a real construction site, wide arrays of input features) and the unfavourable circumstances (mixed mechanised-manual working environment) to prediction performance addressed in this study compared to those available in the literature (e.g. using performance data from manufacturers' handbook; adopting data augmentation techniques and other approaches to generate training data; simplification of the formulation of the prediction problem by fixing some parameters and reducing input features), the prediction performance are considered adequate. This is evidenced with the data in Section 5 which showed that better prediction performance was obtained in days with high excavation volumes when human intervention in the excavation work alongside the excavators is low. The main conclusions from this stage are: (1) while the findings of prediction are adequate, they are more accurate in a highly mechanised approach (i.e. earthwork tasks with equipment mainly and limited human task interventions) compared to a mechanised-manual mixed working environment; and (2) the performance achieved can be improved with a larger training dataset and further parameters testing in future.

The final step in the research explored the use of the DNN prediction to develop benchmark measures that can be used for benchmarking in earthwork projects. An excavation rate was defined as an adequate measure for this purpose. It enables a bottom-up benchmark approach that can be used for benchmarking purposes within the same project and across projects at different granularity levels (e.g.

individual equipment, a set of equipment, work area or zone). For example, when the excavation rate is determined for a work zone according to all vehicle weights involved in the work zone, this method enables comparison between zones, and determines which are operating on-target and which are operating at reduced productivity. If this data can be collected daily (e.g. streamed from existing telematics systems or other IoT systems), it is possible to improve project control by quickly identifying which areas or work zones require more attention, investigating causes of reduced productivity and ensuring that issues are dealt with promptly. More generally, frequent measurement according to the proposed approach is expected to help practitioners identify positive or negative productivity trends; evaluate the impact of productivity improvement techniques (e.g. addition of a new or a different equipment to the fleet); and investigate areas of risks for productivity. These capabilities are not currently available to practitioners with current project control practices within the earthwork sector as identified from the interviews.

In addition to the findings and implications, some general areas of development would benefit the entire approach and unlock the aforementioned capabilities for project control. One such area is the more frequent or even live streaming of data. Recent research by Zohoori et al. (2018) shows that live-streaming of data for production performance monitoring is important where projects are exposed to different kinds of risks and uncertainties such as rework, failure of machines, lack of materials and emergency situations resulting in deviation from plans leading to delay and over-budget delivery of products. However, these applications have generally been applied within controlled settings, i.e. within manufacturing assembly plants within enclosed buildings. Risks and uncertainties within the construction environment are more variable due to the nature of outside working, for example, weather conditions, unknown sub-ground conditions, labour quality and demand, and site logistics (e.g. traffic, emissions regulations). Notably, this case study encountered additional complexities as the site was archaeologically sensitive and required atypical interaction between manual (archaeologists) and mechanised excavation for the removal of human remains during earth excavation. As demonstrated in this paper, these peculiarities added to the challenge of establishing the data set that is required to train the DNN. The data set required to be extended over a 21-day period to capture seasonal trends (i.e. construction equipment performance is dependent on ground conditions as a result of weather), a variety of equipment types (i.e. small plant 1.5 tonne to 9 tonne, to large plant >20 tonne), and varying ground conditions. Telemetry systems offered by OEM are unlikely to fulfil this need for their inconsistency of capture and reporting of data. Future developments should address this issue in two complementary ways: (1) through an IoT system that enable single versions of original data (i.e. no storage of replicated/duplicated data) and a unified approach to data collection across all mixed fleet; and (2) development of an industry standard for OEM telematics systems to remove current inconsistencies in types of data captured, definition of variables, and their reporting.

The first IoT approach requires the delivery of data to a central data storage repository that can ensure an un-edited version of the data remains. Data processing routines such as SQL or Java-based script routines (e.g. Apache Beam) can then provide an auditable and repeatable intervention on the raw data. The application of such processing routines also is more computationally efficient, because data are not being stored in their processed state, rather the data query is stored, and data are processed when needed.

Improved integration and automation of the proposed solution is also required in future work. Integration into a seamless data pipeline can help the stakeholders understand the outputs and value of the proposed solution. Two key areas of improvement include: (1) changing the current process of creating data instances as static files and presenting features to the machine learning model as CSV instances into SQL data storage, and implementing data processing scripts to convert the raw data into a state that is needed for presentation to the AI model (e.g. converting units into aggregations, or data smoothing, removing anomalies, etc.); and (2) adopting a communication service that creates either a continuous stream of data to the data storage or transfers data either as live streamed data or in daily batches depending on how different users interact with the output.

8 Conclusions and implications

The overarching goal of this study was to understand and improve the current process for measuring the productivity of excavators involved in earthwork activities of infrastructure projects. To this end, the paper developed and evaluated a DNN model for measuring the productivity of excavators and benchmarking their performance.

Existing studies focussed on equipment productivity using machine learning approaches were analysed in relation to their similarities and differences with this study, confirming both the gap in knowledge and the novelty of the proposed approach. The interviews with industry experts exposed the current challenges in measuring and controlling of equipment in infrastructure project. The processes for estimating the productivity of sites in terms of volume of earth removed are manual, slow and inaccurate. With current processes, it is also difficult to identify work areas with reduced productivity and guide the controlling function in identifying and monitoring improvement interventions.

The paper developed and tested a DNN model for measuring the volume of earth excavated and benchmarking performances of excavation work. Data sets were obtained from telematics systems over a 21-day period. Feature engineering enabled identification of the relevant input features required for the DNN model. Different gradient descent optimisation algorithms, learning rates, number of layers and hidden nodes were tested to identify the optimal configuration of the DNN.

The prediction performance of the optimal configuration of the DNN was discussed. Due to the lack of studies providing a direct measure of the volume of earth removed as an output of the machine learning algorithm and using field data to both train and predict the output, the comparison of the prediction performance has to be made with studies addressing similar construction applications (e.g. excavators with simulated data, or a tunnel boring machine with real data). The comparative performance of the optimal configuration was evaluated using both the R^2 and the WAPE methods revealing adequate precision levels. Another approach used to judge the precision level achieved was to inspect the details of prediction within the context of the implementation environment. The analysis revealed that more accurate predictions were obtained in days with high excavation volumes whereas precision decreased in days with low excavation volumes, with the reason being the potential involvement of manual work (i.e. archaeologists) alongside some of the equipment which causes gaps in telematics data. This means that the proposed approach and its prediction performance are adequate, but they are more appropriate for a highly mechanised approach (i.e. earthwork tasks with equipment mainly and limited human task interventions) compared to a mixed mechanised-manual working environment.

The paper also demonstrated how the proposed approach can be used to develop benchmarks for earthwork projects using a bottom-up approach spanning across individual equipment or a set of equipment, through work areas, to a whole site. These findings have practical implications to project control as they enable identification of work areas whose productivity is lagging and subsequently inform the resolution of issues causing the productivity gap.

Future work requires an improved integration of data pipelines that can benefit from live streaming, and improved storage and processing. It also requires the development of either a dedicated IoT device or an industry standard for OEM telematics systems to overcome issues of inconsistencies in types of data captured, definitions of variables, and their reporting.

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. and Huntingdon, J. (2021). Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction, Volume 124, 2021, 103532, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103532.

Acknowledgements

This research is supported by Innovate UK (Grant Number: 22724, Project title: Optimising equipment use in construction with BIM, IoT and data analytics) under the 'Increase productivity, performance and quality in UK construction' competition. We owe gratitude to our industry research partners, BIM Academy, BuildStream, and Costain, and the industry experts who provided invaluable input and insightful discussions.

Appendix A: Data used in feature engineering

The following data inputs were given in the OEM telematics system:

- Attachment(hr)
- Auto idle(hr)
- Bibro(hr) i.e. attachment type
- Breaker(hr) i.e. attachment type
- Bucket operation time or others(hr)
- Bucket volume
- CO2 emission(kg)
- Crusher(hr) i.e. attachment type
- Digging time
- Engine on time
- Equipment weight
- Fuel rate
- Fuel remaining(%)
- Fuel used
- Hyd. Oil Temp. Histogram 100 deg. C or above(hr)
- Hyd. Oil Temp. Histogram 50 - 90 deg. C(hr)
- Hyd. Oil Temp. Histogram 90 - 100 deg. C(hr)
- Hyd. Oil Temp. Histogram Under 50 deg. C(hr)
- Latitude
- Longitude
- Not operating
- Operating no dig
- Operating no travel
- Operation (Ex. Travel)(hr)
- Operation time except attachment(hr)
- Radiator Water Temp. Histogram 105 deg. C or above(hr)
- Radiator Water Temp. Histogram 80 - 94 deg. C(hr)
- Radiator Water Temp. Histogram 94 - 105 deg. C(hr)
- Radiator Water Temp. Histogram Under 80 deg. C(hr)
- Ripper(hr) i.e. attachment type
- Swing time
- Travel (Hi)(hr) (i.e. travelling time in high gear)
- Travel (Lo)(hr) (i.e. travelling time in low gear)
- Travelling time

The following were removed for being incomplete or non-numeric

- Attachment(hr)
- Bibro(hr) i.e. attachment type
- Breaker(hr) i.e. attachment type

- Crusher(hr) i.e. attachment type
- Hyd. Oil Temp. Histogram 100 deg. C or above(hr)
- Hyd. Oil Temp. Histogram 90 - 100 deg. C(hr)
- Latitude
- Longitude
- Radiator Water Temp. Histogram 105 deg. C or above(hr)
- Radiator Water Temp. Histogram 94 - 105 deg. C(hr)
- Ripper(hr) i.e. attachment type

The following were removed for improving model convergence

- Auto idle(hr)
- Bucket operation time or others(hr)
- Fuel remaining(%)
- Fuel used
- Hyd. Oil Temp. Histogram 50 - 90 deg. C(hr)
- Hyd. Oil Temp. Histogram Under 50 deg. C(hr)
- Operation (Ex. Travel)(hr)
- Operation time except attachment(hr)
- Radiator Water Temp. Histogram 80 - 94 deg. C(hr)
- Radiator Water Temp. Histogram Under 80 deg. C(hr)
- Travel (Hi)(hr) (i.e. travelling time in high gear)
- Travel (Lo)(hr) (i.e. travelling time in low gear)
- CO2 emission(kg)

The following were retained in the model

- Bucket volume
- Digging
- Engine on time
- Equipment weight
- Fuel rate
- Not operating
- Operating no dig
- Operating no travel
- Swing time
- Total fuel consumed
- Travel time
- Volume excavated

References

Ahn, C., Lee, S. and Peña-Mora, F. (2013). Accelerometer-Based Measurement of Construction Equipment Operating Efficiency for Monitoring Environmental Performance. in *ASCE International Workshop on Computing in Civil Engineering*, Los Angeles, California, United States. https://doi.org/10.1061/9780784413029.071

Ali, I., Sabir, S. and Ullah, Z. (2016). Internet of Things security, device authentication and access control: a review. International Journal of Computer Science and Information Security (IJCSIS), Vol. 14, No. 8. https://arxiv.org/abs/1901.07309v1

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. and Huntingdon, J. (2021). Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction, Volume 124, 2021, 103532, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103532.

Arashpour, M., Ngo, T., & Li, H. (2020). Scene understanding in construction and buildings using image processing methods: A comprehensive review and a case study. Journal of Building Engineering, Vol. 33, 101672. https://doi.org/10.1016/j.jobe.2020.101672

Asadi, A., Alsubaey, M. and Makatsoris, C. (2015). A machine learning approach for predicting delays in construction logistics. International Journal of Advanced Logistics, Vol. 4, issue 2, pp. 115-130. https://doi.org/10.1080/2287108X.2015.1059920

Aslan, B., and Koo, D. (2012). Productivity enhancement for maintenance equipment operations using telematics technology. in Construction Research Congress, West Lafayette, Indiana, United States. https://doi.org/10.1061/9780784412329.098

Azar, E. R. and McCabe, B. (2011). Vision-based equipment detection in construction images, in The 3rd International/9th Construction Specialty Conference, Ottawa, Ontario, Canada, pp. 120–123. https://www.researchgate.net/publication/287024067_Vision-based_equipment_detection_in_construction_images (accessed December 14, 2020).

Baker, H., Hallowell, M.R. and Tixier, A.J.P. (2020). AI-based prediction of independent construction safety outcomes from universal attributes. Automation in Construction, Vol. 118, 103146. https://doi.org/10.1016/j.autcon.2020.103146.

Brownlee, J. (2017a). *Gentle Introduction to the Adam Optimization Algorithm for Deep Learning*, https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/ (accessed December 14, 2020).

Brownlee, J (2017b). A Gentle Introduction to Mini-Batch Gradient Descent and How to Configure Batch Size, https://machinelearningmastery.com/gentle-introduction-mini-batch-gradient-descent-configure-batch-size/ (accessed December 14, 2020).

BSi. (2013). PAS 1192-2:2013 Specification for information management for the capital/delivery phase of construction projects using building information modelling, https://shop.bsigroup.com/forms/PAS-1196-23/ (accessed December 14, 2020).

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

Cheng, M.Y., Kusoemo, D., Gosno, R.A. (2020). Text mining-based construction site accident classification using hybrid supervised machine learning. Automation in Construction, Vol. 118, 103265. https://doi.org/10.1016/j.autcon.2020.103265

Cheng, M.-Y., Tsai, H.-C., & Sudjono, E. (2010). Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. Expert Systems with Applications, Vol. 37, pp. 4224-4231. https://doi.org/10.1016/j.eswa.2009.11.080

Choi, J., Gu, B., Chin, S. and Lee, J.S. (2020). Machine learning predictive model based on national data for fatal accidents of construction workers. Automation in Construction, Vol 110, 102974. https://doi.org/10.1016/j.autcon.2019.102974

Construction Industry Training Board (CITB) (2017). UK Industry Performance Report. https://constructingexcellence.org.uk/wp-content/uploads/2018/11/UK-Industry-Performance-Report-2017.pdf (accessed December 14, 2020).

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. and Huntingdon, J. (2021). Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction, Volume 124, 2021, 103532, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103532.

Dixit, M., Tiwari, A., Pathak, H. and Astya, R. (2018). An overview of deep learning architectures libraries and its applications areas, in 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida (UP), India, pp. 293-297. https://doi.org/10.1109/ICACCCN.2018.8748442

Edwards, D.J. and Holt, G.D. (2000), "ESTIVATE: a model for calculating excavator productivity and output costs", Engineering, Construction and Architectural Management, Vol. 7 No. 1, pp. 52-62. https://doi.org/10.1108/eb021132

Edwards, D. J. and Griffiths, I. J. (2000) Artificial intelligence approach to calculation of hydraulic excavator cycle time and output. Mining Technology: Transactions of the Institutions of Mining and Metallurgy: Section A, Vol. 109, Issue 1, pp. 23-29. https://doi.org/10.1179/mnt.2000.109.1.23

Egan, J. (1998). Rethinking Construction. 1st Ed. HMSO, London. Available at: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf (accessed December 14, 2020)

European Commission, Directorate-General for Regional and Urban Policy (2014). Guide to Cost-Benefit Analysis of Investment Projects, https://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/cba_guide.pdf (accessed December 14, 2020). doi:10.2776/97516
Fahmy, A., Altaf, H., Al Nabulsi, A., Al-Ali, A. and Aburukba, R. (2019). Role of RFID Technology in Smart City Applications, in 2019 International Conference on Communications, Signal Processing, and their Applications (ICCSPA), Sharjah, United Arab Emirates, pp. 1-6. https://doi.10.1109/ICCSPA.2019.8713622

Fang, Y., Cho, Y., Zhang, S. and Perez, E. (2016). Case Study of BIM and Cloud–Enabled Real-Time RFID Indoor Localization for Construction Management Applications, *Journal of Construction Engineering and Management*, Vo. 142, Issue 7. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001125

Farmer, M. (2016). The Farmer Review of the UK Construction Labour Model. Construction Leadership Council (CLC). https://www.constructionleadershipcouncil.co.uk/wp-content/uploads/2016/10/Farmer-Review.pdf (accessed December 14, 2020).

Gartner (2019). Gartner says 5.8 billion enterprise and automotive IOT endpoints will be in use in 2020. https://www.gartner.com/en/newsroom/press-releases/2019-08-29-gartner-says-5-8-billion-enterprise-and-automotive-io (accessed December 14, 2020).

Ghosh, A., Hosseini, M. R., Edwards, D., Kassem, M., Matteo-Garcia, M. (2019). Use cases for the Internet of Things (IoT) in the construction sector: Lessons from leading industries. Proceedings of the 36th International Conference of CIB W78, Newcastle-upon-Tyne, UK, 18-20 September, pp. 882-889 (ISSN: 2706-6568) http://itc.scix.net/paper/w78-2019-paper-083

Gondia, A., Siam, A., El- Dakhakhni, W. and Nasar, A.H. (2020). Machine Learning Algorithms for Construction Projects Delay Risk Prediction, Journal of Construction Engineering and Management, Vol 146 Issue 1. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001736

Greater London Authority (2014). The control of dust and emissions during construction and demolition supplementary planning guidance, https://www.london.gov.uk/sites/default/files/gla_migrate_files_destination/Dust%20and%20Emissions%20SPG%208%20July%202014.pdf (accessed December 14, 2020).

Guest, G., Bunce, A. and Johnson, L. (2006). How many interviews are enough? an experiment with data saturation and variability. Field Methods, Vol. 18, No. 1, pp. 59-82. https://doi.org/10.1177%2F1525822X05279903

Hackitt, D. J. (2018). Building a Safer Future. UK Secretary of State for Housing, Communities and Local Government ISBN 978-1-5286-0293-8 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/707785/Building_a_Safer_Future_-_web.pdf (accessed December 14, 2020).

Institute of Practitioners in Advertising (IPA) (2019). Best Practice in Benchmarking. https://www.gov.uk/government/publications/best-practice-in-benchmarking (accessed December 14, 2020).

ISO (2020). ISO/TS 15143-3:2020 Earth-moving machinery and mobile road construction machinery - Worksite data exchange - Part 3: Telematics data. https://www.iso.org/standard/76394.html (accessed December 14, 2020).

Jagushte, V.J. (2017). Usability Review of Telematics for Construction Equipment Fleet Management. Master Thesis, University of Florida. https://ufdc.ufl.edu/UFE0051120/00001 (accessed December 14, 2020).

Jia, X., Feng, Q., Fan, T. and Lei, Q. (2012). RFID technology and its applications in Internet of Things (IoT), 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), Yichang China, 21-23 April 2012, pp. 1282-1285. https://doi.10.1109/CECNet.2012.6201508

Kanan, R., Elhassan, O. and Bensalem, R. (2018). An IoT-based autonomous system for workers' safety in construction sites with real-time alarming, monitoring, and positioning strategies, *Automation in Construction*, 88, 73-86. https://doi.org/10.1016/j.autcon.2017.12.033

Kassem, M., Li, J., Kumar, B., Malleson, A., Gibbs, D., Kelly, G. & Watson, R. (2020). Building Information Modelling: Evaluating Tools for Maturity and Benefits Measurement, Centre for Digital Built Britain: Cambridge, UK. https://www.cdbb.cam.ac.uk/news/bim-report-evaluating-tools-maturity-and-benefits-measurement (accessed December 14, 2020).

Kassem, M., Rogage, K., Huntingdon, J., Durojaye, G., Arena, N., Kelly, G, Lund, T., and Clarke, T. (2019). Measuring and Improving the Productivity of Construction's Site Equipment Fleet: An integrated IoT and BIM System. Proceedings of the 36th International Conference of CIB W78, Newcastle-upon-Tyne, UK, 18-20 September, pp. 901-911, ISSN: 2706-6568. http://itc.scix.net/paper/w78-2019-paper-085

Kim C., Kim, B., and Kim, H. (2013). 4D CAD model updating using image processing-based construction progress monitoring. Automation in Construction, 35, 44-52. https://doi.org/10.1016/j.autcon.2012.11.041

KPMG (2015). UK Construction Barometer: Margins Under Pressure. 1st Ed. KPMG, London. https://app.croneri.co.uk/whats-new/construction-industry-still-not-balance (accessed December 14, 2020).

Latham, M. (1994). Constructing the Team. ISBN 0 11 752994 X https://constructingexcellence.org.uk/wp-content/uploads/2014/10/Constructing-the-team-The-Latham-Report.pdf (accessed December 14, 2020).

Lu, W., Huang, G.Q. and Li, H. (2011). Scenarios for applying RFID technology in construction project management, *Automation in Construction*,2, 101-106, 2011. https://doi.org/10.1016/j.autcon.2010.09.007

Makridakis, S. (1994). Accuracy measures: theoretical and practical concerns. International Journal of forecasting, Vol. 9 Issue 4, pp. 527-529. https://doi.org/10.1016/0169-2070(93)90079-3

Marzouk, M. and Zaher, M. (2020). Artificial intelligence exploitation in facility management using deep learning, Construction Innovation, Vol. 20 Issue 4. https://doi.org/10.1108/CI-12-2019-0138

McKINSEY (2017). Reinventing Construction: a route to higher productivity. McKinsey Global Institute, https://www.mckinsey.com/mgi/overview (accessed December 14, 2020).

Memarzadeh, M., Golparvar-Fard, M., and Niebles, J.C. (2013). Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors, Automation in Construction, Vol 32, pp. 24-37. https://doi.org/10.1016/j.autcon.2012.12.002

Nagpal, A. (2017). L1 and L2 Regularization Methods, Towards Data Science, https://towardsdatascience.com/l1-and-l2-regularization-methods-ce25e7fc831c (accessed December 14, 2020).

Nath, N.D., Behzadan, A.H., Paal, S.G. (2020). Deep learning for site safety: Real-time detection of personal protective equipment, Automation in Construction, Vol. 112. https://doi.org/10.1016/j.autcon.2020.103085.

Ngo, N. (2018). Early predicting cooling loads for energy-efficient design in office buildings by machine learning, Energy and Buildings, Vol. 182, pp. 264-273. https://doi.org/10.1016/j.enbuild.2018.10.004.

Niu, Y., Lu, W., Liu, D., Chen, K. and Xue, F. (2017). A Smart Construction Object (SCO)-enabled proactive data management system for construction equipment management, in ASCE International Workshop on Computing in Civil Engineering 2017, pp. 130-138. https://doi.org/10.1061/9780784480830.017

Poh, C. Q.X, Ubeynarayana, C.U., and Goh, Y.M. (2018). Safety leading indicators for construction sites: A machine learning approach, Automation in Construction, Vol. 93, pp. 375-386. https://doi.org/10.1016/j.autcon.2018.03.022.

Rashid, K.M. and Louis, J., 2019. Times-series data augmentation and deep learning for construction equipment activity recognition. Advanced Engineering Informatics, Vol. 42. https://doi.org/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, Vol 119. https://doi.org/10.1016/j.autcon.2020.103361

Rehman, A. (2019). Why AEMP 2.0 is Critical to Managing Your Mixed Fleet. https://blog.orbcomm.com/why-aemp-2-0-is-critical-to-managing-your-mixed-fleet/ (accessed December 14, 2020).

Romdhani, S. (2015). Implementation of DNN-HMM acoustic models for phoneme recognition, PhD Dissertation, University of Waterloo http://hdl.handle. net/10012/9061 (accessed December 14, 2020).

Ruder, S. (2016). *An overview of gradient descent optimization algorithms*, https://ruder.io/optimizing-gradient-descent/index.html#adagrad (accessed December 14, 2020).

Said, H., Nicoletti, T., and Perez-Hernandez, P. (2014). Utilizing telematics data to support effective equipment fleet-management decisions: utilization rate and hazard functions. *Journal of Computing in Civil Engineering*, Vol. 30 Issue 1. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000444

Saunders, M., Lewis, P. & Thornhill, A. (2015). *Research methods for business students, 7th Ed*, Pearson Education, https://www.pearson.com/uk/educators/higher-education-educators/program/Saunders-Research-Methods-for-Business-Students-7th-Edition/PGM1089011.html?tab=resources (accessed December 14, 2020).

Schabowicz, K. and Hola, B. (2007). Mathematical-neural model for assessing productivity of earthmoving machinery, Journal of Civil Engineering and Management, Vol. 13 Issue 1, pp. 47-54. https://doi.org/10.1080/13923730.2007.9636418

Seung C. Ok & Sunil K. Sinha (2006) Construction equipment productivity estimation using artificial neural network model, Construction Management and Economics, Vol. 24 Issue 10, pp. 1029-1044. https://doi.org/10.1080/01446190600851033

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. and Huntingdon, J. (2021). Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction, Volume 124, 2021, 103532, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103532.

Singaravel, S., Suykens, J., and Geyer, P. (2018). Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction, Advanced Engineering Informatics, Vol 38, 2018, pp. 81-90. https://doi.org/10.1016/j.aei.2018.06.004.

Succar, B., and Kassem, M., (2015). Macro-BIM adoption: conceptual structures. Automation in Construction, Vol. 57, pp. 64-79. https://doi.org/10.1016/j.autcon.2015.04.018

Sulankivi, K., Zhang, S., Teizer, J., Eastman, C. M., Kiviniemi, M., Romo, I., & Granholm, L. (2013). Utilization of BIM-based Automated Safety Checking in Construction Planning. in Proceedings of the 19th CIB World Building Congress, Brisbane Australia, pp. 5-9. https://external-apps.qut.edu.au/wbc2013/papers/cibwbc2013_submission_268.pdf

Tajeen, H., Zhu, Z. (2014). Image dataset development for measuring construction equipment recognition performance, Automation in Construction, Vol. 48, pp. 1-10. https://doi.org/10.1016/j.autcon.2014.07.006

Tam, C., Tong, T.K. and Tse, S.L. (2002). Artificial neural networks model for predicting excavator productivity. Engineering Construction and Architectural Management, Vol. 9, pp. 446-452. https://doi.org/10.1108/eb021238

Tatari, O., & Kucukvar, M. (2011). Cost premium prediction of certified green buildings: A neural network approach. Building and Environment Vol. 46 Issue 5, pp.1081-1086. https://doi.org/10.1016/j.buildenv.2010.11.009

Tian, J., Gurley, K.R., Diaz, M.T., Fernández-Cabán, P.L., Masters, F.J. and Fang, R., (2020). Low-rise gable roof buildings pressure prediction using deep neural networks. Journal of Wind Engineering and Industrial Aerodynamics, Vol. 196. https://doi.org/10.1016/j.jweia.2019.104026

Tixier, A.J.P., Hallowell, M.R., Rajagopalan, B. and Bowman, D. (2016). Application of machine learning to construction injury prediction, Automation in Construction, Vol. 69, pp. 102-114. https://doi.org/10.1016/j.autcon.2016.05.016.

Thomas A.J., Petridis M., Walters S.D., Gheytassi S.M., Morgan R.E. (2017). Two hidden layers are usually better than one. in Proceedings of 18th International Conference of Engineering Applications of Neural Networks (EANN) 2017. Athens, Greece, August 25–27, 2017. pp. 279-290. https://doi.org/10.1007/978-3-319-65172-9_24

Wolstenholme, A. (2009). Never Waste a Good Crisis. 1st Ed. Constructing Excellence, London. Available at: https://constructingexcellence.org.uk/wp-content/uploads/2014/12/Wolstenholme_Report_Oct_2009.pdf (accessed December 14, 2020).

Wu, Y., Kim, H., Kim, C., & Han, S. H. (2009). Object recognition in construction-site images using 3D CAD-based filtering. *Journal of Computing in Civil Engineering*, 24(1), 56-64. https://doi.org/10.1061/(ASCE)0887-3801(2010)24:1(56)

Yu, W. and Skibniewski, M.J. (2010). Integrating Neurofuzzy System with Conceptual Cost Estimation to Discover Cost-Related Knowledge from Residential Construction Projects, Journal of Computing in Civil Engineering, Vol 24, Issue 1, pp. 35–44. https://doi.org/10.1061/(ASCE)0887-3801(2010)24:1(35)

Zankoula, E., Khoury, H., Salhaba, D., and Antara S. (2018). A review on Internet of Things solutions for enhancing construction equipment fleet productivity. in *Proceedings of the Creative Construction Conference* 2018, 30 June - 3 July 2018, Ljubljana, Slovenia, pp. 971 – 978. http://dx.doi.org/10.3311/CCC2018-126

Zhong, D., LV, H., Han, J. and Wei, Q. (2014). A Practical Application Combining Wireless Sensor Networks and Internet of Things: Safety Management System for Tower Crane Groups, *Sensors*, Vol. 14, Issue 8, pp. 13794-

Kassem, M., Mahamedi, E., Rogage, K., Duffy, K. and Huntingdon, J. (2021). Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction, Volume 124, 2021, 103532, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103532.

13814. https://doi.org/10.3390/s140813794

Zhou, C. and Ding, L. (2017). Safety barrier warning system for underground construction sites using Internet-of-Things technologies, *Automation in Construction*, Vol. 83, pp. 372-389. https://doi.org/10.1016/j.autcon.2017.07.005

Zohoori, B., Verbraeck, A., Bagherpour, M. and Khakdaman, M. (2018). Monitoring production time and cost performance by combining earned value analysis and adaptive fuzzy control, *Computers and Industrial Engineering,* Vol. 127, pp. 805-821. https://doi.org/10.1016/j.cie.2018.11.019