




Leveraging convolutional neural networks for efficient classification of heavy construction equipment

Mohamed S. Yamany¹ · Mohamed M. Elbaz² · Ahmed Abdelaty³ · Mohamed T. Elnabwy^{4,5} 

Received: 12 August 2024 / Accepted: 16 August 2024
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract

Effective classification and detection of equipment on construction sites is critical for efficient equipment management. Despite substantial research efforts in this field, most previous studies have focused on classifying a limited number of equipment categories. Furthermore, there is a scarcity of research dedicated to heavy construction equipment. Hence, this study develops a robust Convolutional Neural Network (CNN) model to classify heavy construction machinery into 12 different types. The study utilizes a comprehensive dataset of equipment images, which was divided into three distinct subsets: 60% for training the model, 30% for validating its performance, and 10% for testing its accuracy. The model's robustness was ensured by monitoring accuracy and loss measures during the training and validation phases. The CNN model achieved approximately 85% training accuracy with a minimum loss of 0.40. The testing phase revealed a high overall precision of 80%. The CNN model accurately classifies concrete mixer machines and telescopic handlers with an Area Under the Curve (AUC) of 0.92, however pile driving machines have a lower accuracy with an AUC of 0.83. These findings demonstrate the model's high ability to distinguish between several types of heavy construction equipment. This paper contributes to the relatively unexplored area of classifying heavy construction equipment by providing a practical tool for automating equipment classification, leading to enhanced efficiency, safety, and maintenance protocols in construction management.

Keywords Heavy construction equipment · Classification · Convolutional neural network · Deep learning

✉ Mohamed S. Yamany
msyamany@eng.zu.edu.eg; mohamedyamany9@gmail.com

Mohamed M. Elbaz
mobaz@mc.edu.eg

Ahmed Abdelaty
aahmed3@uwoy.edu

Mohamed T. Elnabwy
mohamed_elnabwy@nwr.gov.eg;
Eng.elnabwy@yahoo.com

- ¹ Department of Construction Engineering, Faculty of Engineering, Zagazig University, Zagazig 44519, Egypt
- ² Mansoura College Academy, Mansoura - Damietta High Way, El Mansoura, Egypt
- ³ Department of Civil and Architectural Engineering and Construction Management, University of Wyoming Laramie, Laramie WY82071, USA
- ⁴ Coastal Research Institute (CORI), National Water Research Center, Alexandria 21415, Egypt
- ⁵ Civil Engineering Department, Faculty of Engineering, Damietta University, New Damietta 34517, Egypt

Introduction

Highway construction projects rely heavily on the efficient management and deployment of a wide range of heavy machinery. Equipment such as excavators, which are used for ground preparation, and dump trucks, which handle material transportation, is critical to the successful completion of various construction phases (Kim, Kim et al., 2018b; Nath & Behzadan, 2020). Traditionally, equipment classification and identification were based on manual inspection by trained personnel (Akhavian & Behzadan, 2015; Cheng et al., 2010). Although this method can achieve a certain level of accuracy, it has several limitations. The manual classification is inherently time-consuming and resource-intensive. Furthermore, the potential for human error can cause inconsistencies and inaccuracies, particularly in large-scale projects involving a variety of equipment types (Sherafat et al., 2020).

Deep Learning (DL) methods, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks

(RNNs), have emerged as powerful tools in a variety of fields, including construction management (Elghaish et al., 2022; Ji et al., 2023; Park et al., 2023; Soltani et al., 2016; Yabuki et al., 2018). Unlike artificial neural networks (ANNs), the capability of CNN models to autonomously identify complex patterns and representations from raw data offers promising opportunities for optimizing construction workflows and improving equipment management. Nevertheless, there is a research gap in the literature regarding the use of DL techniques - especially CNNs - to classify heavy machinery employed in highway construction projects (Akinosho et al., 2020).

Prior research has predominantly concentrated on applying DL to predict equipment failures, optimize maintenance schedules, and monitor construction progress. These studies highlight the potential of DL to transform multiple facets of construction management (Bunrit et al., 2019; Jung et al., 2022). However, there has been limited exploration into the specific challenges of classifying heavy equipment, particularly in highway construction projects (Akinosho et al., 2020). Additionally, existing research focuses on classifying a limited number of equipment categories, lacking comprehensive coverage of the diverse array of heavy machinery used in these projects. To bridge these research gaps, this paper aims at achieving the following research objectives:

1. Conduct a review of existing literature to identify previous studies on the classification of heavy construction equipment in highway construction projects to establish the novelty and significance of the study.
2. Evaluate previous studies on classifying construction equipment to understand existing approaches and limitations.
3. Develop a CNN model to classify a wide range of heavy construction equipment in highway projects, addressing limitations in previous studies focusing on fewer equipment classes.
4. Rigorously test the CNN model to demonstrate its accurate classification of heavy construction equipment, validating its potential for real-world applications.

Literature review

Precise classification and detection of construction equipment is crucial for enhancing project efficiency, ensuring safety, and optimizing resource allocation (Kim et al., 2018a). This allows for more efficient resource allocation, minimizing maintenance expenses and project delays (Post et al., 2018; Slaton et al., 2020a). By effectively monitoring equipment on construction sites, construction managers

can improve productivity, reduce downtime, and mitigate risks (Mohy et al., 2024; Xu et al., 2023; Yan et al., 2017). Ultimately, real-time equipment monitoring contributes to keeping projects on track and within budget.

Traditional classification techniques like ANNs, Support Vector Machines (SVMs), and k-Nearest Neighbors (kNN) have been used in various classification tasks (Anirudh et al., 2023; Elshaboury et al., 2024; Kaveh, 2024a, b; Kaveh & Khavaninzadeh, 2023; Obianyo et al., 2023; Yamany, 2020; Zihan et al., 2023). These algorithms depend largely on manually extracted features that are created and fed into the algorithm. However, these classification models are limited by their learning capabilities and heavy reliance on expert domain knowledge to define features (Akinosho et al., 2020; Fang et al., 2016; Li et al., 2023). In contrast, the advent of DL, particularly CNNs, has transformed the field. CNNs have the capability to automatically learn relevant features directly from raw image data, eliminating the need for manual feature extraction (Xiao & Kang, 2021; Zhao et al., 2020). Groundbreaking research has been conducted on the use of CNNs in equipment classification, demonstrating that even shallow CNN architectures can be effective in tasks such as monitoring excavators. For example, one study found that CNNs could classify seven different excavator activities with 90.7% accuracy using data from inertial measurement unit signals (Slaton et al., 2020a). This success is attributed to CNN's capability to efficiently extract spatial features from sensor data using parallel convolution operations.

Over the last decade, the use of DL for detecting construction equipment has expanded substantially. Table 1 provides a comprehensive comparison of various DL-based recognition techniques, serving as a valuable source for understanding the current landscape of DL applications in construction. This table outlines research efforts across different sub-fields of the broader construction domain, highlighting the versatility and growing importance of DL in addressing the challenges associated with the detection of construction equipment.

There has been little emphasis in the literature on the classification and detection of heavy equipment used in highway construction projects. For example, Arabi et al. (2020) developed a practical DL approach for detecting six types of construction equipment used in highway construction. This approach achieved a mean average precision of over 90%, making it suitable for real-time construction applications such as safety monitoring and productivity assessment. In addition to classification and detection of construction equipment, other studies have investigated various aspects of equipment usage, including productivity for modular construction safety, which used R-CNN and achieved a precision of 0.890 (Zheng et al., 2020). Wang et

Table 1 Summary of studies used DL for object classification and detection on construction sites

No.	Reference	Application	DL Task	No. Classes	DL Method	Accuracy	Precision	Recall	F1 Score
1	(Rashid & Louis, 2019)	Excavator equipment monitoring and classification	Classification	10	LSTM	0.979	0.962	0.990	0.976
2	(Guo et al., 2020)	Safety	Detection	6	CNN	N/A	0.988	N/A	N/A
3	(Slaton et al., 2020b)	Monitoring activities of heavy construction equipment	Tracking compactor activity	6	CNN LSTM	0.744 0.752	0.750 0.780	0.730 0.760	0.720 0.750
5	(Shi et al., 2020)	Safety	Classification	4	LSTM	High	N/A	N/A	N/A
6	(Akhavian & Behzadan, 2015)	Monitoring equipment activity	Monitoring	5	ANN, KNN, DT	0.985	N/A	N/A	N/A
7	(Xiao et al., 2021)	Monitoring construction machines at night-time	Tracking	10	Gladnet	0.951	0.759	0.963	N/A
8	(Xiao & Kang, 2021)	Monitoring activities from a safety perspective	Detection	10	Inception-SSD R-FCN-ResNet101	N/A	0.830 0.888	N/A	N/A
9	(Lu et al., 2021)	Productivity	Detection	11	RESNET with faster R-CNN	0.952	0.926	N/A	N/A
10	(Lu et al., 2021)	Productivity	Detection	11	RESNET with faster R-CNN	0.952	0.926	N/A	N/A
11	(Jung et al., 2022)	Safety and productivity	Detection	13	CNN	N/A	0.821	0.831	0.826
12	(Jung et al., 2023)	Monitoring	Detection	19	CNN	0.885	0.869	0.840	0.854
13	(Wang et al., 2023)	Safety	Detection	6	GhostNet	N/A	0.921	0.903	N/A
14	(Sharma & Sen, 2020)	Monitoring	Detection	12	CNN	0.726–0.990	N/A	N/A	N/A
15	(Zheng et al., 2020)	Module detection for modular construction safety	Detection	13	R-CNN	N/A	0.890	0.830	N/A
16	(Wang et al., 2022)	Monitoring construction sites for safety purposes	Detection	12	DeepLabV3+	0.926	N/A	N/A	N/A
17	(Braun et al., 2020)	Monitoring	Detection	11	CNN	N/A	0.914	0.927	0.920
18	(Shen et al., 2024)	Monitoring equipment activities	Detection	8	TCN	0.945	0.944	0.945	0.944
19	(Ding et al., 2018)	Detecting unsafe behaviour for safety purposes	Detection	8	CNN, LSTM	0.970	N/A	N/A	N/A
20	(Hernandez et al., 2019b)	Monitoring	Detection	6	CNN, LSTM	0.771	N/A	N/A	N/A
21	(Arabi et al., 2020)	Safety monitoring, productivity assessments	Detection	6	CNN	0.900	N/A	N/A	N/A

al. (2022) developed a DeepLabV3+ model for monitoring construction sites with an accuracy of 0.926, whereas Braun et al. (2020) created a CNN model for monitoring construction tasks with a recall of 0.914 and an F1 score of 0.927. Moreover, Xiao and Kang (2020) focused on productivity-related tasks, illustrating the potential of DL techniques to optimize equipment utilization and operational efficiency. Furthermore, Shen et al. (2024) applied a Temporal Convolutional Network (TCN) model for monitoring equipment activities, achieving precision and recall scores of 0.945 and 0.944, respectively.

Most DL models developed for classifying and detecting construction equipment address a limited number of classes. Studies such as Ding et al. (2018); Hernandez et al. (2019a) focused on fewer than ten equipment classes. Ding et al. (2018) achieved a high accuracy of 0.970 in detecting unsafe behaviour using a CNN model, while Hernandez et

al. (2019a) obtained an accuracy of 0.771 for general monitoring of equipment activity tasks using an LSTM model. In contrast, few studies have developed models for more than ten classes. For instance, Shen et al. (2024) and Nath et al. (2020) explored classification tasks involving a higher number of equipment categories, highlighting the need for further research in this area to develop more robust models capable of handling a broader range of equipment types. According to the literature review conducted in this study, most prior studies have concentrated on detecting and classifying construction equipment into ten or fewer classes. This underscores the necessity for advancements in DL models to handle more comprehensive classifications, particularly in complex and dynamic construction environments.

Overview of CNN model

Object classification and detection technology has evolved significantly, transitioning from methods that relied on hand-crafted features like Scale-Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) to deep learning approaches, particularly CNN (Nath et al., 2020). In contrast to traditional ANNs, which use image pixels directly for classification, CNN models simplify this process by consolidating weights into smaller kernel filters, which enhances learning efficiency and robustness. CNN represents a powerful type of deep neural network capable of directly learning complex patterns from data, leading to substantial advancements in object detection, image classification, speech recognition, and feature extraction (Fang et al., 2018; Huang et al., 2018; Zhang, 2022). CNN networks are structured with foundational components that enable sophisticated image processing, which are as follows:

- **Convolutional Layers:** These layers apply convolution operations to input images, producing feature maps emphasizing specific visual patterns. During training, the network identifies and prioritizes important features necessary for accurate image scanning and categorization, as depicted in Fig. 1. The convolution operation can be mathematically represented as

$$\text{Output}[i, j] = \sum (\text{Filter}[m, n] \times \text{Input}[i + m, j + n]) + \text{Bias} \quad (1)$$

Where $\text{Output}[i, j]$ represents the value located at position (i, j) within the feature map; $\text{Filter}[m, n]$ denotes the value positioned at (m, n) within the filter; $\text{Input}[i + m, j + n]$

corresponds to the value found at position $(i + m, j + n)$ within the input image; and Bias is a trainable value that adjusts the output of the filter.

- **Pooling Layers:** Situated between convolutional layers, pooling layers down sample feature maps while retaining essential features extracted by preceding layers. A common pooling operation is max pooling, which selects the maximum value within each pooling window.
- **Activation Functions:** Non-linear activation functions like ReLU introduce non-linearity into the network, enhancing its ability to learn complex relationships. The ReLU function is expressed as.

$$\text{ReLU}(x) = \max(0, x) \quad (2)$$

- **Fully Connected Layers:** At the final stages, fully connected layers process flattened outputs from convolutional layers to compute class probabilities using the SoftMax activation function for classification.

Research methodology

Figure 2 illustrates the systematic methodology employed for heavy construction equipment image classification using a CNN model.

Construction equipment image data collection

A comprehensive dataset of heavy construction equipment images was meticulously constructed for training and evaluating the CNN model. This dataset encompasses

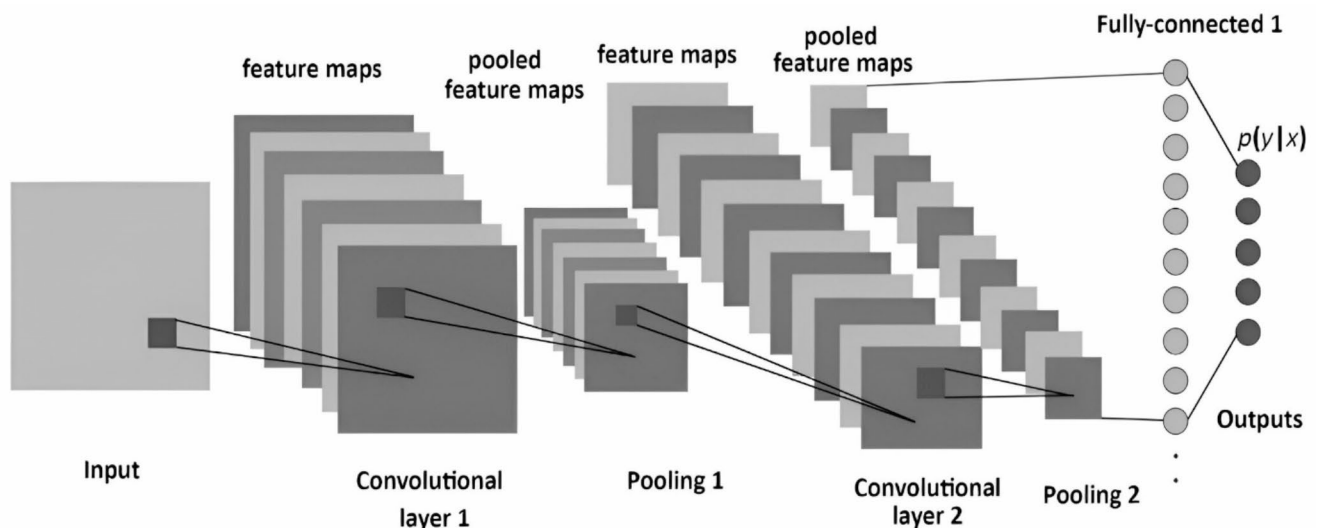


Fig. 1 Convolution operation in CNN model (Albelwi & Mahmood, 2017)

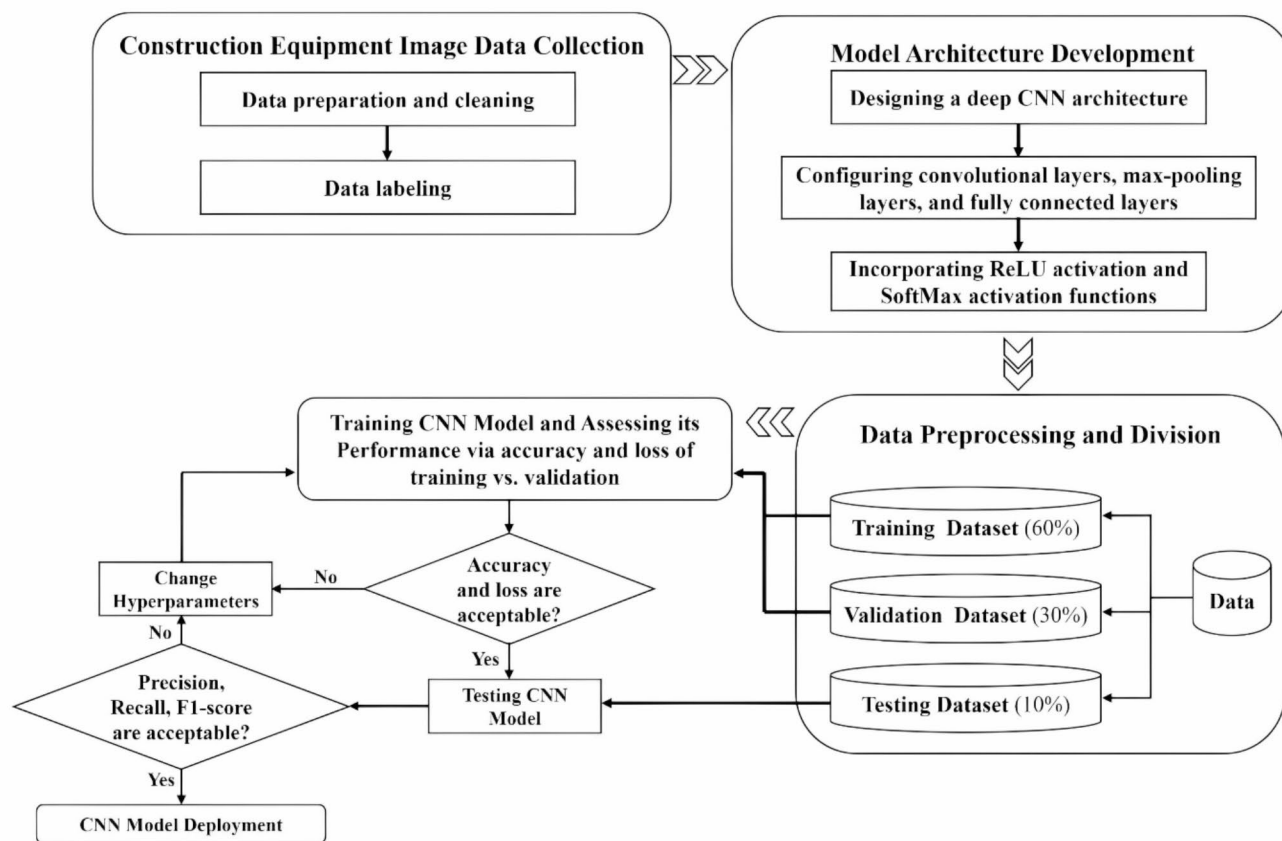


Fig. 2 Research methodology

10,846 images categorized into 12 distinct classes, ensuring a diverse representation of various equipment types (e.g., excavators and loaders). The dataset was divided into three subsets:

1. Training Dataset (60%, 6,595 images): This subset was used to train the CNN model, allowing it to learn the complex relationships between image features and their respective equipment classes.
2. Validation Dataset (30%, 3,291 images): This subset helped monitor the model's performance throughout training to mitigate the risk of overfitting. High performance on the validation set indicates the model's ability to generalize to new, unseen data.
3. Testing Dataset (10%, 960 images): This subset was used for the final evaluation after the training phase, providing an unbiased measure of the model's accuracy in real-world classification tasks.

To maintain consistency and facilitate interpretation, each image was assigned a unique numeric identifier ranging from 0 to 11, corresponding to its specific equipment class. This labeling system simplifies the referencing and analysis of classification results. By using sequentially organized

labels, the model's predictions can be easily matched with the respective equipment types, improving the clarity and understanding of the results for subsequent data analysis, comparisons, and decision-making processes. Table 2 provides the labels and descriptions for the 12 equipment classes.

CNN model architecture development

To handle the complexities of construction equipment images, we designed a deep CNN architecture inspired by established models like VGG and ResNet (Simonyan & Zisserman, 2015). This architecture leverages multiple convolutional layers for feature extraction. Each convolutional layer uses rectified linear unit (ReLU) activations to introduce non-linearity and improve model performance. Max-pooling layers are strategically inserted between convolutional layers to reduce image dimensionality while preserving key features. The model's depth is carefully chosen to capture intricate visual details crucial for distinguishing between various construction equipment types. As illustrated in Fig. 3, the network follows a sequential structure:

Table 2 Equipment code, label, and description

Code	Label	Description
0	Asphalt roller	A heavy machine used to compact and smooth asphalt pavement after it is laid.
1	Boom lift	A mobile platform with an extendable and articulated arm that can elevate workers and materials to various heights for construction, maintenance, or other tasks.
2	Concrete mixer machine	A machine used to combine cement, sand, gravel, and water to produce concrete.
3	Concrete mixer truck	A truck equipped with a rotating drum that mixes concrete ingredients while transporting them to the construction site.
4	Dump truck	A large, open-bed truck used to haul loose materials such as dirt, gravel, sand, or demolition debris.
5	Excavator	A tracked or wheeled machine with a long arm and bucket used for digging, trenching, loading, and demolition.
6	Forklift	A powered industrial vehicle with a two-pronged fork at the front for lifting and transporting pallets and other materials.
7	Loader	A wheeled machine with a large bucket at the front for loading, scooping, and transporting loose materials.
8	Motor grader	A machine with a long, adjustable blade used for grading, leveling, and shaping surfaces such as roads, driveways, and building sites.
9	Pile driving machine	A specialized machine used to drive piles (long, strong poles) into the ground to provide support for foundations, bridges, and other structures.
10	Scissor lift	A mobile platform that rises vertically on scissor-like mechanisms, providing temporary access for workers and materials at various heights.
11	Telescopic Handler	A machine with a telescopic boom that can be extended and retracted horizontally. It can be equipped with various attachments for lifting, placing, and transporting materials.

1. Convolutional Layers: The process starts with a convolutional layer containing 16 filters of size 3×3 . This layer extracts low-level features from the input image. Subsequent convolutional layers, with increasing numbers of filters (e.g., 32), progressively extract more complex features.
2. Max-Pooling Layers: Interspersed between convolutional layers are max-pooling layers. These layers reduce the image size while retaining the most relevant features extracted by the preceding convolutional layers.
3. Fully Connected Layers: After feature extraction, the process transitions to fully connected layers. The flattened output from the final max-pooling layer is fed into a fully connected layer with 256 neurons and ReLU activation. This layer performs non-linear transformations

on the extracted features. Finally, a second fully connected layer with a number of neurons equal to the equipment categories is employed. This layer utilizes the SoftMax activation function to generate probabilities for each equipment class, enabling multi-class classification.

CNN model training and validation

The CNN model was trained using the training dataset, while the validation set was used to evaluate the model's performance throughout the training process. An appropriate optimizer (Adam) and a categorical cross-entropy loss function were utilized to reduce the classification error (Liu et al., 2023). During the training process, accuracy and loss metrics were continuously monitored for both the training and validation datasets. These metrics guided iterative adjustments to the model. Successful training is shown by high performance on the validation dataset; if not, the model architecture was refined. Refinements could include adding batch normalization, dropout layers, or other architectural changes. To prevent overfitting, techniques such as early stopping were employed to halt training when validation accuracy stopped improving or started declining. Additionally, data augmentation was used to artificially increase the size of the training dataset, providing the model with a wider range of examples for each class and thus mitigating overfitting. In instances of underfitting, where the model failed to capture the complexity of the data, the model's capacity was increased, typically by adding more convolutional layers or neurons. Ultimately, the training process was conducted using a Jupyter Notebook, optimized for performance on a system equipped with an Intel (R) Core (TM) i7-10510U CPU @ 1.80 GHz, boosting up to 2.30 GHz.

CNN model testing and evaluation metrics

After successfully completing the training and validation processes, the final CNN model underwent rigorous testing using a separate testing dataset that had not been seen during training or validation. This test dataset was utilized to evaluate the model's performance in accurately classifying heavy construction equipment images. The effectiveness of the CNN model in real-world scenarios was thoroughly assessed by analyzing performance metrics, including precision, recall, and F1-score.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (3)$$

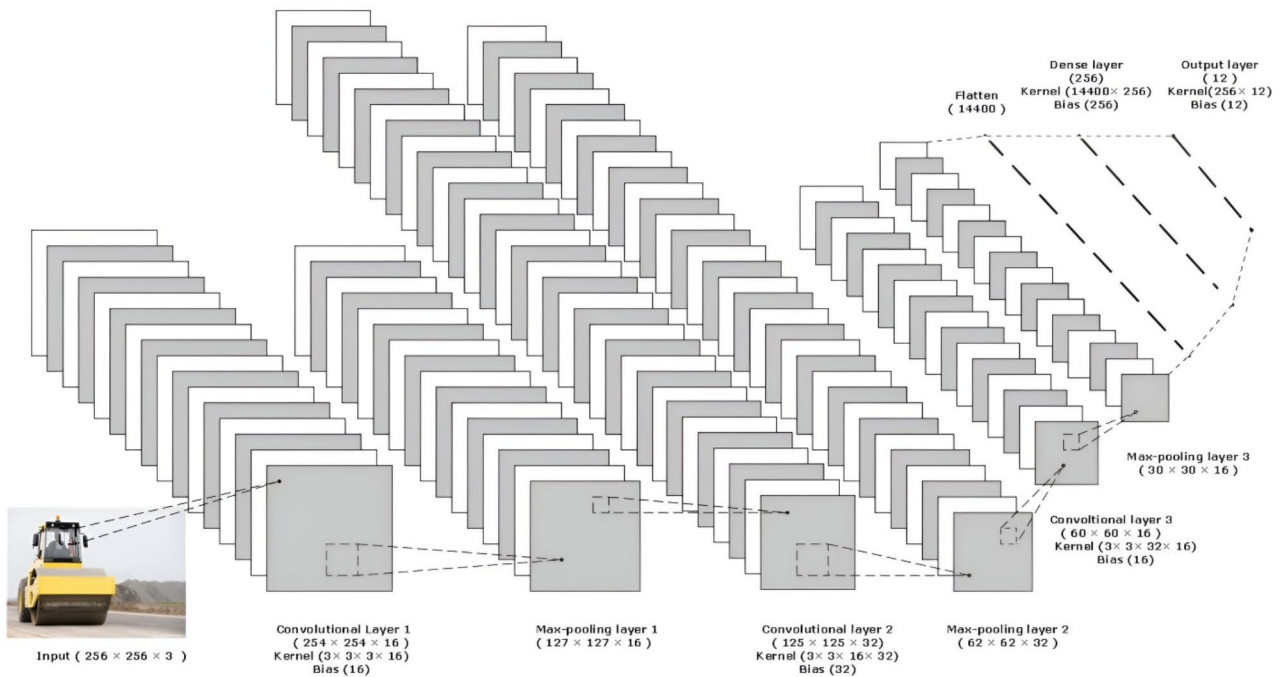


Fig. 3 Detailed architecture of the CNN model

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{4}$$

$$\text{F1} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

Results and discussion

Various CNN architectures with different configurations and hyperparameters were explored during the training phase, and this section discusses the results of the training, validation and testing of the optimal design.

Performance evaluation of CNN model during training stage

Figure 4 shows the accuracy and loss curves of training and validation. The training accuracy curve shows a steady and continuous increase, reflecting effective learning and classification of the training data. The validation accuracy curve also progresses positively, indicating that the model generalizes well to new images. Additionally, the training loss curve, which consistently declines, indicates the model’s

successful adaptation to reduce errors. The validation loss curve similarly decreases, suggesting the model’s capability to generalize and make accurate predictions on the validation data. The minimal variation in loss during training, along with its steady convergence to a low value (0.4), implies that the optimizer effectively finds the global minimum of the loss function. Overall, the training and validation curves reveal that the model has effectively learned the complex features of heavy construction equipment, achieving commendable accuracy and low loss metrics on both datasets. These results highlight the model’s potential to accurately classify and identify different types of construction equipment, contributing to enhanced operational efficiency, maintenance, and safety in the construction industry.

Performance evaluation of CNN model during testing stage

The classification results presented in Table 3 offer a detailed assessment of the model’s effectiveness in categorizing heavy construction equipment into 12 distinct classes during the testing phase. The precision scores average around 0.80, with a range from 0.71 to 0.87. Notably, the model shows high precision in categories like concrete mixer trucks (0.87), boom lifts (0.86), and telescopic

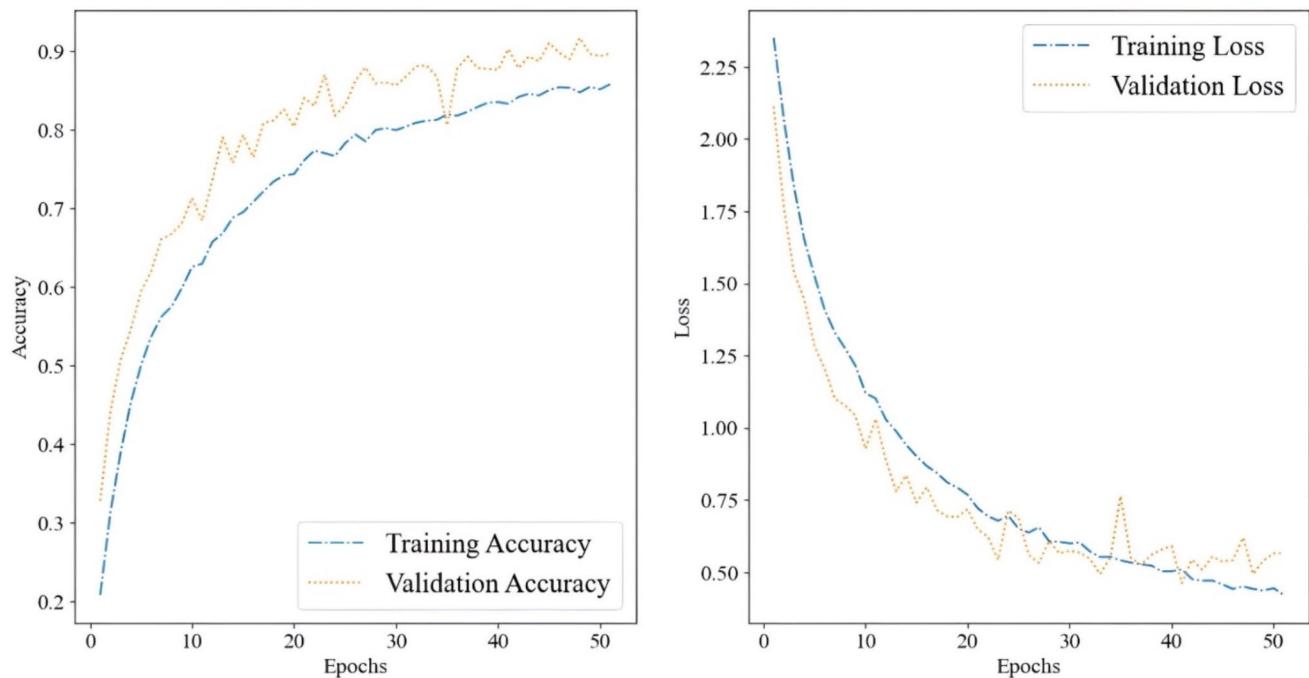


Fig. 4 Training accuracy and loss vs. validation accuracy and loss

Table 3 Classification report of testing phase

Class Code	Class Name	Precision	Recall	F1-score	Support
0	Asphalt roller	0.79	0.81	0.80	79
1	Boom lift	0.86	0.86	0.86	80
2	Concrete mixer machine	0.80	0.77	0.78	82
3	Concrete mixer truck	0.87	0.83	0.85	72
4	Dump truck	0.81	0.76	0.78	82
5	Excavator	0.78	0.80	0.79	66
6	Forklift	0.78	0.78	0.78	92
7	Loader	0.71	0.79	0.75	68
8	Motor grader	0.77	0.84	0.80	83
9	Pile driving machine	0.81	0.73	0.77	83
10	Scissor lift	0.79	0.79	0.79	85
11	Telescopic handler	0.84	0.83	0.83	88

handlers (0.84), indicating its high accuracy in identifying these specific equipment types. The variation in precision scores might be due to differences in visual complexity and distinctiveness among the classes, with equipment having more easily identifiable features achieving higher precision. Moreover, the recall scores, which reflect the model's ability to correctly identify all relevant instances within a class, range from 0.73 to 0.86. The highest recall score of 0.86 was observed for Class 1 (boom lift), demonstrating the model's high capability to detect true positives in this category. Conversely, the lower recall rate of 0.73 for Class 9 (pile driving machine) could be due to visual similarities with other

equipment types or challenges in correctly identifying all instances.

Furthermore, the F1-score, which balances precision and recall, ranges from 0.75 to 0.86. Class 1 (boom lift) attained the highest F1-score of 0.86, while Class 7 (loader) had the lowest F1-score of 0.75. The lower score for Class 7 suggests an imbalance between precision and recall, possibly due to challenges in accurately distinguishing this class based on visual features alone. Furthermore, the support values, indicating the number of instances per class, range from 66 to 92. Classes with higher support generally have more training data, which may contribute to better classification performance.

Overall, these metrics indicate that the model performs competitively in classifying heavy construction equipment. However, certain challenges persist, particularly in classes with lower precision and recall. Addressing these issues may require refining the model's feature extraction capabilities and enhancing the training process to improve accuracy and generalization across all equipment categories.

To comprehensively assess the performance of the CNN model, the Receiver Operating Characteristic (ROC) curve, a metric for assessing classification model performance, was created and investigated. Figure 5 displays the ROC curves for the 12 distinct types of construction equipment. The model exhibits impressive performance, as evidenced by its high Area Under the Curve (AUC) values for all classes. Notably, the model achieves an AUC score of 0.92 for both the concrete mixer machine and telescopic handler (classes

Fig. 5 ROC curves of equipment classification for testing stage

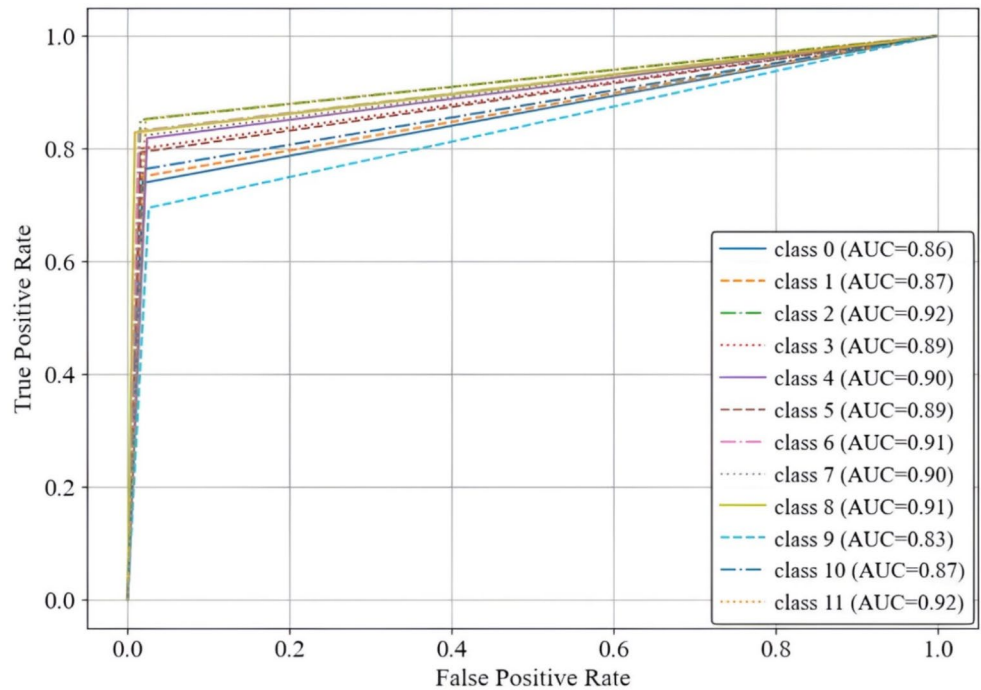
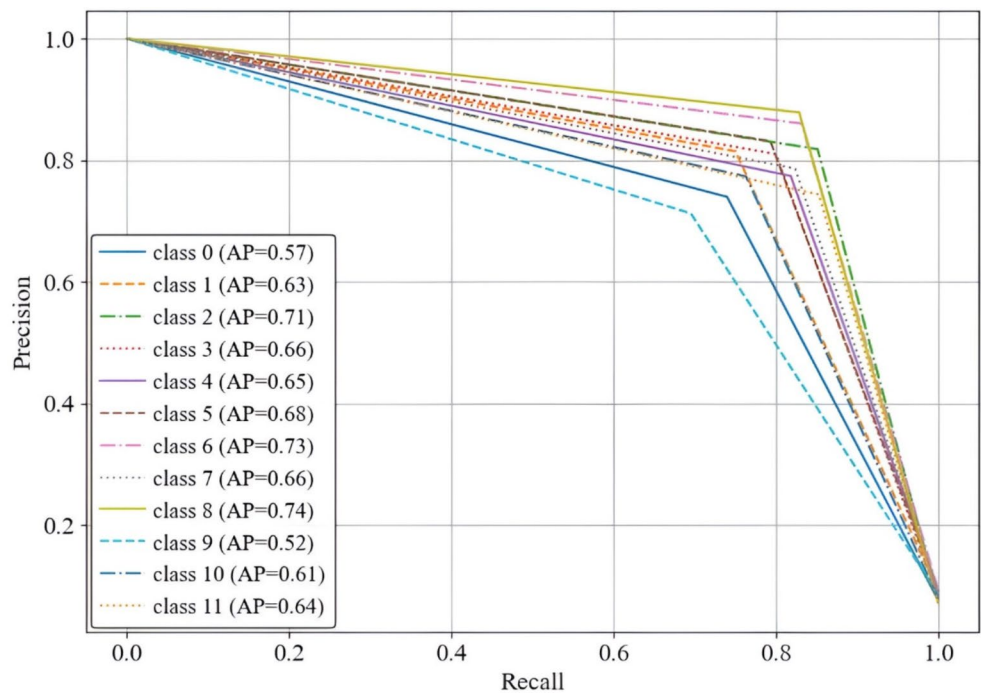


Fig. 6 Precision-recall curves of equipment classification for testing stage



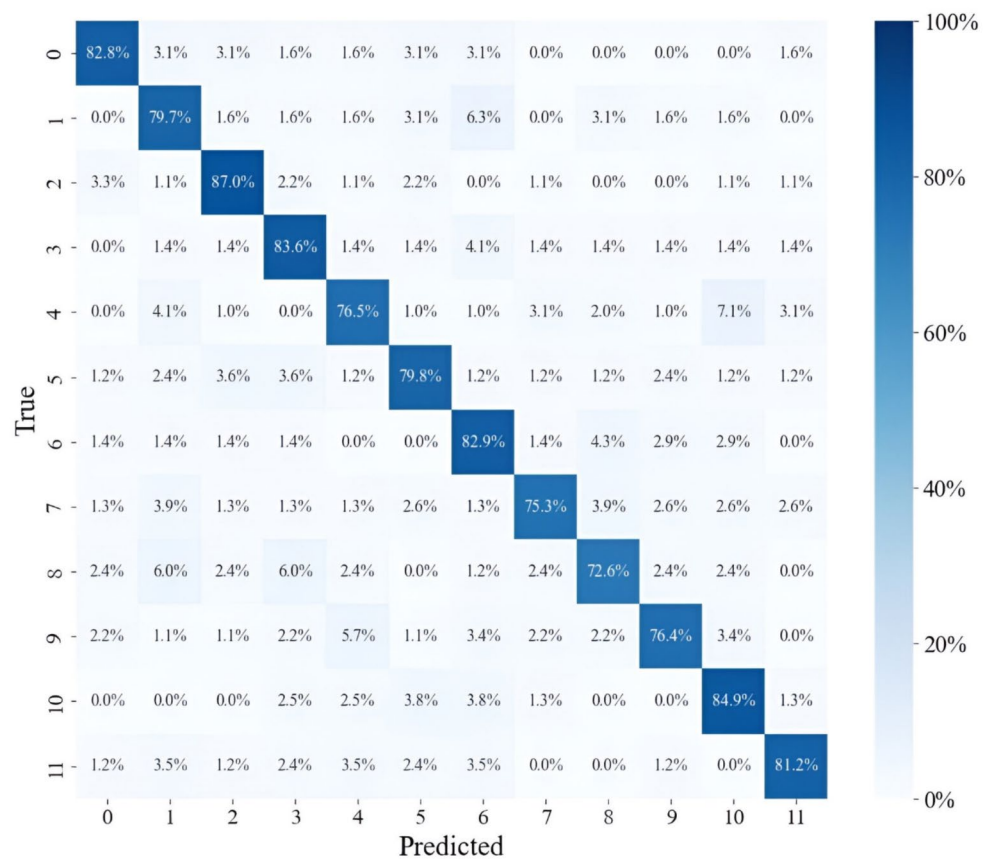
2 and 11), indicating highly accurate classification. Moreover, both the forklift (class 6) and motor grader (class 8) demonstrate high performance, with an AUC value of 0.91.

However, the pile driving machine (class 9) exhibits a lower AUC value of 0.83, indicating difficulties in accurately classifying this particular equipment type. The graphical representation in Fig. 5 provides a visual overview of the AUC metrics across different construction equipment categories, offering insights into the classifier’s effectiveness

in distinguishing between equipment types based on the testing results. Besides, the precision-recall curves in Fig. 6 support these findings, showing that classes 6 and 8 achieve the best results, with average precision (AP) values of 0.73 and 0.74, respectively. In contrast, class 9 records the lowest precision-recall performance, with an AP score of 0.52.

In a separate effort, a confusion matrix was constructed during the testing phase to evaluate the CNN model’s performance in classifying construction equipment (Fig. 7).

Fig. 7 Confusion matrix of equipment classification for testing stage



The matrix shows that the model achieved high accuracy in classifying concrete mixer machines, scissor lifts, concrete mixer trucks, and forklifts (classes 2, 10, 3, and 6, respectively). However, there is a room for improvement in accurately classifying asphalt rollers, telescopic handlers, excavators, and boom lifts (classes 0, 11, 5, and 1, respectively). To enhance the model's performance, applying data augmentation techniques and fine-tuning the model's hyperparameters could be beneficial. Additionally, a more in-depth analysis of the misclassified images, especially those with lower accuracy, may provide valuable insights for improving the model's ability to distinguish between specific types of equipment.

Conclusions

This paper introduces a CNN model developed specifically to tackle the challenge of accurately classifying heavy construction equipment on construction sites. This model represents a significant advancement in the identification and categorization of various heavy equipment in the construction industry. The developed CNN model demonstrates remarkable accuracy, with precision scores ranging from 0.71 to 0.87 and recall values ranging from 0.73 to 0.86

across various equipment classes. These findings highlight the model's effectiveness in accurately distinguishing different types of construction machinery.

The real-world implications of adopting this CNN model are substantial. The model contributes to optimizing operational efficiency and logistics in construction projects by automating the identification and categorization of equipment on construction site. This results in enhanced resource allocation and more efficient equipment tracking, which leads to improved project execution. Furthermore, the CNN model enhances safety protocols on construction sites by providing a robust system for detailed equipment monitoring. This facilitates improved recognition of potential hazard and focused actions for risk mitigation, fostering a safer working environment for construction personnel. Additionally, the model's streamlined operations and improved maintenance practices contribute to cost reductions.

It is crucial to acknowledge that the developed CNN model has some limitations. The model classifies a specific set of equipment on construction site. Moreover, there exists a potential imbalance in the training dataset due to data limitations. However, we can unlock the model's full potential by addressing these limitations through future research efforts. Promising avenues for future exploration include integrating real-time data streams for continuous monitoring and

adaptation, utilizing transfer learning techniques to expand applicability to a broader range of equipment categories, and investigating advanced image augmentation techniques to mitigate potential dataset biases and improve the model's overall robustness.

Author contributions The authors confirm their contribution to the paper as follows: study conception and design: M.S. Yamany, M.M. Elbaz; analysis and interpretation of results: M.S. Yamany, M.M. Elbaz, A. Abdelaty, M.T. Elnabwy; draft manuscript preparation: M.S. Yamany, M.M. Elbaz, A. Abdelaty, M.T. Elnabwy; All authors reviewed the results and approved the final version of the manuscript.

Data availability Data employed in this research study is available upon request from the corresponding author.

Declarations

Competing interests The authors declare no competing interests.

References

- Akhavian, R., & Behzadan, A. H. (2015). Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers. *Advanced Engineering Informatics*, 29(4), 867–877. <https://doi.org/10.1016/J.AEI.2015.03.001>
- Akinosho, T. D., Oyedele, L. O., Bilal, M., Ajayi, A. O., Delgado, M. D., Akinade, O. O., & Ahmed, A. A. (2020). Deep learning in the construction industry: A review of present status and future innovations. *Journal of Building Engineering*, 32, 101827. <https://doi.org/10.1016/J.JOBE.2020.101827>
- Albelwi, S., & Mahmood, A. (2017). A framework for designing the architectures of deep convolutional neural networks. *Entropy* 2017, 19(6), 242. <https://doi.org/10.3390/E19060242>. 19.
- Anirudh, N., Padala, S. P. S., & Reddy, H. N. E. (2023). Development of ANN-Based Risk Prediction Model in Construction Projects. In K. R. Reddy, S. Kalia, S. Tangellapalli, & D. Prakash (Eds.), *Recent Advances in Sustainable Environment* (pp. 109–121). Springer Nature. https://doi.org/10.1007/978-981-19-5077-3_9
- Arabi, S., Haghghat, A., & Sharma, A. (2020). A deep-learning-based computer vision solution for construction vehicle detection. *Computer-Aided Civil and Infrastructure Engineering*, 35(7), 753–767. <https://doi.org/10.1111/MICE.12530>
- Braun, A., Tuttas, S., Borrmann, A., & Stilla, U. (2020). Improving progress monitoring by fusing point clouds, semantic data and computer vision. *Automation in Construction*, 116. <https://doi.org/10.1016/J.AUTCON.2020.103210>
- Bunrit, S., Kerdprasop, N., & Kerdprasop, K. (2019). Evaluating on the transfer learning of CNN architectures to a construction material image classification task. *International Journal of Machine Learning and Computing*, 9(2), 201–207. <https://doi.org/10.18178/ijmlc.2019.9.2.787>
- 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*, 37(6), 4224–4231. <https://doi.org/10.1016/J.ESWA.2009.11.080>
- Ding, L., Fang, W., Luo, H., Love, P. E. D., Zhong, B., & Ouyang, X. (2018). A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Automation in Construction*, 86, 118–124. <https://doi.org/10.1016/J.AUTCON.2017.11.002>
- Elghaish, F., Matarneh, S. T., & Alhusban, M. (2022). The application of deep learning in construction site management: Scientometric, thematic and critical analysis. *Construction Innovation*, 22(3), 580–603. <https://doi.org/10.1108/CI-10-2021-0195/FULL/PDF>
- Elshaboury, N., Yamany, M. S., Labi, S., & Smadi, O. (2024). Enhancing local road pavement condition prediction using bayesian-optimized ensemble machine learning and adaptive synthetic sampling technique. *International Journal of Pavement Engineering*, 25(1), 2365957. <https://doi.org/10.1080/10298436.2024.2365957>
- Fang, Y., Cho, Y. K., Zhang, S., & 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*, 142(7). [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001125](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001125)
- Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., & Li, C. (2018). Computer vision aided inspection on falling prevention measures for steeplejacks in an aerial environment. *Automation in Construction*, 93, 148–164. <https://doi.org/10.1016/j.autcon.2018.05.022>
- Guo, Y., Xu, Y., & Li, S. (2020). Dense construction vehicle detection based on orientation-aware feature fusion convolutional neural network. *Automation in Construction*, 112, 103124. <https://doi.org/10.1016/J.AUTCON.2020.103124>
- Hernandez, C., Slaton, T., Balali, V., & Akhavian, R. (2019a). A deep learning framework for construction equipment activity analysis. *Computing in Civil Engineering 2019: Data, Sensing, and Analytics - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2019*, 479–486. <https://doi.org/10.1061/9780784482438.061>
- Hernandez, C., Slaton, T., Balali, V., & Akhavian, R. (2019b). A deep learning framework for construction equipment activity analysis. *Computing in Civil Engineering 2019: Data, Sensing, and Analytics - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2019*, 479–486. <https://doi.org/10.1061/9780784482438.061>
- Huang, L., Li, J., Hao, H., & Li, X. (2018). Micro-seismic event detection and location in underground mines by using Convolutional neural networks (CNN) and deep learning. *Tunnelling and Underground Space Technology*, 81, 265–276. <https://doi.org/10.1016/j.tust.2018.07.006>
- Ji, A., Xue, X., Zhang, L., Luo, X., & Man, Q. (2023). A transformer-based deep learning method for automatic pixel-level crack detection and feature quantification. *Engineering, Construction and Architectural Management, ahead-of-print(ahead-of-print)*. <https://doi.org/10.1108/ECAM-06-2023-0613/FULL/PDF>
- Jung, S., Jeoung, J., Kang, H., & Hong, T. (2022). 3D convolutional neural network-based one-stage model for real-time action detection in video of construction equipment. *Computer-Aided Civil and Infrastructure Engineering*, 37(1), 126–142. <https://doi.org/10.1111/MICE.12695>
- Jung, S., Jeoung, J., Lee, D. E., Jang, H., & Hong, T. (2023). Visual-auditory learning network for construction equipment action detection. *Computer-Aided Civil and Infrastructure Engineering*, 38(14), 1916–1934. <https://doi.org/10.1111/MICE.12983>
- Kaveh, A. (2024a). Artificial intelligence: Background, applications and future. In A. Kaveh (Ed.), *Applications of artificial neural networks and machine learning in civil engineering* (pp. 1–53). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-66051-1_1
- Kaveh, A. (2024b). Buckling Resistance Prediction of High-Strength Steel Columns Using Metaheuristic-Trained Artificial Neural Networks. In A. Kaveh (Ed.), *Applications of Artificial Neural Networks and Machine Learning in Civil*

- Engineering* (pp. 55–73). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-66051-1_2
- Kaveh, A., & Khavaninzadeh, N. (2023). Efficient training of two ANNs using four meta-heuristic algorithms for predicting the FRP strength. *Structures*, 52, 256–272. <https://doi.org/10.1016/j.istruc.2023.03.178>
- Kim, H., Kim, H., Hong, Y. W., & Byun, H. (2018b). Detecting construction equipment using a region-based fully convolutional network and transfer learning. *Journal of Computing in Civil Engineering*, 32(2). [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000731](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000731)
- Kim, H., Bang, S., Jeong, H., Ham, Y., & Kim, H. (2018a). Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation. *Automation in Construction*, 92, 188–198. <https://doi.org/10.1016/J.AUTCON.2018.04.002>
- Li, L., Sun, Q., Wang, Y., & Gao, Y. (2023). A data-driven indirect approach for predicting the response of existing structures induced by adjacent excavation. *Applied Sciences (Switzerland)*, 13(6). <https://doi.org/10.3390/AP13063826>
- Liu, H., Wang, D., Xu, K., Zhou, P., & Zhou, D. (2023). Lightweight convolutional neural network for counting densely piled steel bars. *Automation in Construction*, 146, 104692. <https://doi.org/10.1016/J.AUTCON.2022.104692>
- Lu, J., Yao, Z., Bi, Q., & Li, X. (2021). A neural network-based approach for fill factor estimation and bucket detection on construction vehicles. *Computer-Aided Civil and Infrastructure Engineering*, 36(12), 1600–1618. <https://doi.org/10.1111/MICE.12675>
- Mohy, A. A., Bassioni, H. A., Elgendy, E. O., & Hassan, T. M. (2024). Innovations in safety management for construction sites: The role of deep learning and computer vision techniques. *Construction Innovation, ahead-of-print*(ahead-of-print). <https://doi.org/10.1108/CI-04-2023-0062/FULL/PDF>
- Nath, N. D., & Behzadan, A. H. (2020). Deep Convolutional networks for construction object detection under different visual conditions. *Frontiers in Built Environment*, 6, 532607. <https://doi.org/10.3389/FBUIL.2020.00097/BIBTEX>
- 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*, 112. <https://doi.org/10.1016/J.AUTCON.2020.103085>
- Obianyo, J. I., Udeala, R. C., & Alaneme, G. U. (2023). Application of neural networks and neuro-fuzzy models in construction scheduling. *Scientific Reports*, 13(1), 8199. <https://doi.org/10.1038/s41598-023-35445-5>
- Park, S. M., Lee, J. H., & Kang, L. S. (2023). A framework for improving object recognition of structural components in construction site photos using deep learning approaches. *KSCE Journal of Civil Engineering*, 27(1), 1–12. <https://doi.org/10.1007/S12205-022-2318-0/METRCS>
- Post, V. E. A., Banks, E., & Brunke, M. (2018). Groundwater flow in the transition zone between freshwater and saltwater: A field-based study and analysis of measurement errors. *Hydrogeology Journal*, 26(6), 1821–1838. <https://doi.org/10.1007/S10040-018-1725-2>
- Rashid, K. M., & Louis, J. (2019). Times-series data augmentation and deep learning for construction equipment activity recognition. *Advanced Engineering Informatics*, 42, 100944. <https://doi.org/10.1016/J.AEI.2019.100944>
- Sharma, S., & Sen, S. (2020). One-dimensional convolutional neural network-based damage detection in structural joints. *Journal of Civil Structural Health Monitoring*, 10(5), 1057–1072. <https://doi.org/10.1007/S13349-020-00434-Z>
- Shen, Y., Wang, J., Feng, C., & Wang, Q. (2024). Dual attention-based deep learning for construction equipment activity recognition considering transition activities and imbalanced dataset. *Automation in Construction*, 160, 105300. <https://doi.org/10.1016/J.AUTCON.2024.105300>
- Sherafat, B., Ahn, C. R., Akhavian, R., Behzadan, A. H., Golparvar-Fard, M., Kim, H., Lee, Y. C., Rashidi, A., & Azar, E. R. (2020). Automated methods for activity recognition of construction workers and equipment: State-of-the-art review. *Journal of Construction Engineering and Management*, 146(6), 03120002. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001843](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001843)
- Shi, J., Sun, D., Hu, M., Liu, S., Kan, Y., Chen, R., & Ma, K. (2020). Prediction of brake pedal aperture for automatic wheel loader based on deep learning. *Automation in Construction*, 119, 103313. <https://doi.org/10.1016/J.AUTCON.2020.103313>
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.
- Slaton, T., Hernandez, C., & Akhavian, R. (2020a). Construction activity recognition with convolutional recurrent networks. *Automation in Construction*, 113, 103138. <https://doi.org/10.1016/J.AUTCON.2020.103138>
- Slaton, T., Hernandez, C., & Akhavian, R. (2020b). Construction activity recognition with convolutional recurrent networks. *Automation in Construction*, 113. <https://doi.org/10.1016/J.AUTCON.2020.103138>
- Soltani, M. M., Zhu, Z., & Hammad, A. (2016). Automated annotation for visual recognition of construction resources using synthetic images. *Automation in Construction*, 62, 14–23. <https://doi.org/10.1016/J.AUTCON.2015.10.002>
- Wang, Z., Zhang, Y., Mosalam, K. M., Gao, Y., & Huang, S. L. (2022). Deep semantic segmentation for visual understanding on construction sites. *Computer-Aided Civil and Infrastructure Engineering*, 37(2), 145–162. <https://doi.org/10.1111/MICE.12701>
- Wang, L., Wang, B., Zhang, J., Ma, H., Luo, P., & Yin, T. (2023). An Intelligent Detection Method for Approach Distances of Large Construction Equipment in substations. *Electronics* 2023, 12(16), 3510. <https://doi.org/10.3390/ELECTRONICS12163510.12>
- Xiao, B., & Kang, S. C. (2020). Vision-based method integrating deep learning detection for Tracking multiple construction machines. *Journal of Computing in Civil Engineering*, 35(2), 04020071. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000957](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000957)
- Xiao, B., & Kang, S. C. (2021). Development of an image data set of construction machines for deep learning object detection. *Journal of Computing in Civil Engineering*, 35(2). [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000945](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000945)
- Xiao, B., Lin, Q., & Chen, Y. (2021). A vision-based method for automatic tracking of construction machines at nighttime based on deep learning illumination enhancement. *Automation in Construction*, 127, 103721. <https://doi.org/10.1016/J.AUTCON.2021.103721>
- Xu, N., Liang, Y., Guo, C., Meng, B., Zhou, X., Hu, Y., & Zhang, B. (2023). Entity recognition in the field of coal mine construction safety based on a pre-training language model. *Engineering Construction and Architectural Management, ahead-of-print*(ahead-of-print). <https://doi.org/10.1108/ECAM-05-2023-0512/FULL/PDF>
- Yabuki, N., Nishimura, N., & Fukuda, T. (2018). Automatic object detection from digital images by deep learning with transfer learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10863 LNCS, 3–15. https://doi.org/10.1007/978-3-319-91635-4_1
- Yamany, M. S. (2020). *Stochastic Performance and Maintenance Optimization Models for Pavement Infrastructure Management* [PhD Thesis, Purdue University Graduate School]. https://hammer.purdue.edu/articles/thesis/Stochastic_Performance_and_Maintenance_Optimization_Models_for_Pavement_Infrastructure_Management/12252716

- Yan, X., Li, H., Li, A. R., & Zhang, H. (2017). Wearable IMU-based real-time motion warning system for construction workers' musculoskeletal disorders prevention. *Automation in Construction*, 74, 2–11. <https://doi.org/10.1016/j.autcon.2016.11.007>
- Zhang, F. (2022). A hybrid structured deep neural network with Word2Vec for construction accident causes classification. *International Journal of Construction Management*, 22(6), 1120–1140. <https://doi.org/10.1080/15623599.2019.1683692>
- Zhao, Y., Deng, X., & Lai, H. (2020). A YOLO-Based method to recognize structural components from 2D drawings. *Construction Research Congress 2020: Computer Applications - Selected Papers from the Construction Research Congress 2020*, 753–762. <https://doi.org/10.1061/9780784482865.080>
- Zheng, Z., Zhang, Z., & Pan, W. (2020). Virtual prototyping- and transfer learning-enabled module detection for modular integrated construction. *Automation in Construction*, 120. <https://doi.org/10.1016/J.AUTCON.2020.103387>
- Zihan, Z. U. A., Smadi, O., Tilberg, M., & Yamany, M. S. (2023). Synthesizing the performance of deep learning in vision-based pavement distress detection. *Innovative Infrastructure Solutions*, 8(11), 299. <https://doi.org/10.1007/s41062-023-01250-2>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.