



INVESTIGATION OF AUTOMATION OPPORTUNITIES IN WAREHOUSE MANAGEMENT IN CONSTRUCTION SUPPLY CHAINS USING CONVOLUTIONAL NEURAL NETWORKS

*A Dissanayake ^a, R Sugathadasa ^a and M M De Silva ^{a b *}*

^a Department of Transport Management and Logistics Engineering,
Faculty of Engineering, University of Moratuwa, Katubedda 10400, Sri Lanka.

^b Extreme Energy-Density Research Institute, Nagaoka University of Technology,
Nagaoka, Niigata 940-2188, Japan.

* Correspondence should be addressed to mavinds@uom.lk

ABSTRACT

The building industry is strongly reliant on materials, which account for 55%-60% of its expenses. However, due to outdated and time-consuming approaches, inadequate inventory management prevails. This is where developing technologies such as Deep Learning (DL) might help uncover solutions. Surprisingly, very little scientific research on DL has been conducted for this purpose. As a result, this study looks into the prospect of automating construction warehouse management by employing CNN for object detection and counting. During the initial investigation, 23 studies out of 26 used Convolutional Neural Networks (CNN) for image processing and object detection. Secondly, a model was developed and compared its accuracy to that of human counting and discovered that the model outperformed people. Industry professionals were interviewed to discuss the findings. Industry professionals highlighted the advantages and disadvantages of using such an automated system in construction warehouses. In conclusion, this study shows that the CNN base model outperforms people in counting materials, and the proposed automated inventory management system has significant industry potential.

Keywords: *Inventory Management, Warehouse Automation, Image Processing, Convolutional Neural Networks, Construction Supply Chains*

1. INTRODUCTION

The construction industry products are unique, longer time to produce, stationery products, significantly longer useful lifetime, complex design, and involvement of complex supply chains [1]. Various stakeholders are involved in construction projects throughout the production phase and during the useful lifetime [2]. A significant distinction between construction and manufacturing is that the construction industry is project-based and discontinuous in nature, while the manufacturing industry involves continuous processes and relationships [3]. Construction supply chains are complex due to the involvement of different partners, suppliers, the use of large quantities, and various materials [4]. The definition of the Construction Supply Chain (CSC), according to Tserng et al. [5] is “not a chain of construction businesses with business-to-business relationships, but a network of multiple organisations and relationships, which includes the flow of information, the flow of materials, services or products, and the flow of funds between client, designer, contractor, and supplier” also they mention “the integration of key construction business processes, from the demands of the client, design to construction, and key members of the construction supply chain, including client/owner, designer, contractor, subcontractor and supplier”. CSC has a make-to-order type supply chain model where the product is not made until customer needs arise.

Deep Learning (DL) is an emerging AI technique that is a subset of Machine Learning (ML) that can distinguish the features or the characteristics of the data set without human intervention to define features [6]. Also, DL can improve the model's performance, while older ML techniques stagnate their performance at some point. DL architectures comprise several components; layers, neurons, activation functions, and weights [7]. DL works in a similar mechanism to how human neurons begin working, where data is passed through neurons. Once the data triggers the activation function through weighted connections of the neuron, the neuron will process the data and pass the output to the next neuron/s. Research done by Ruiz-del-Solar et al. [8] suggests several ways that DL can use robotics in areas like; object detection, objects touching & moving, scene representation and classification, sensing the environment, and spatiotemporal vision. Zhao et al. [9] discuss how DL can be used for computer vision tasks like salient object detection, face detection, and pedestrian detection.

Convolutional Neural Networks (CNN) is another architecture under DL and follows the ANN characteristics [6]. Inside a CNN, when an image is input to the network, it breaks into small pieces based on the number of layers, and in a CNN, it is possible to break into every pixel. Convolutional layers are extracted from the input image and find the relationships in input pixels by analysing image properties like; the gradient, curves, shapes, texture, colour, edges, contours, strokes, and orientation. When the

output of convolutional layers comes to the pooling layer, it reduces the dimensionality of the image outputs of convolutional layers. Finally, connecting layers provide the desired output [6]. CNN is commonly used for image processing, image classification, object detection, and face recognition [10]. CNN neurons are three-dimensional width, height, and depth, where the data is stored as a three-dimensional matrix [10]. CNN has different class variations like LeNet, AlexNet, VGGNet, GoogleNet, ResNet, and ZFNet. ResNet is frequently used for deep CNN since it can overcome vanishing gradient issues over other conventional CNN classes [6].

There are already a few attempts to apply new technologies in inventory management, which have proved that new technologies can significantly improve inventory management of construction supply chains. If the construction industry can embrace new technologies like DL in inventory management systems and enhance the industry's productivity, that would be better.

The purpose of this research is to provide DL base suggestions and solutions for the warehouse management construction industry to overcome the challenges they face due to traditional inventory management systems and improve the productivity of their projects. This study aims to (a) propose an automated warehouse solution for the construction industry, (b) find the most suitable DL architecture for object detection and counting, (c) assess the accuracy of DL-based object detection and counting over humans, and (d) identify potential issues and challenges during and after the implementation of the automated system. The development of the object detection and face detection model is presented in Section 3, followed by its calibration and estimates, summarised in Section 4, while the conclusions are discussed in Section 5.

2. LITERATURE REVIEW

The construction cost of materials accounts for 55%-60% of the project's total cost, which means in CSC, proper inventory management is crucial to control the project's costs [11]. Even though inventory management is critical and hectic, the construction industry uses traditional methods like ABC for inventory management, planning, and categorisation [11]. Some surveys about the construction industry have revealed that; a lack of real-time data about stock levels of warehouses, late or early placement of the order for materials, lack of stock management systems, lack of stock planning, lack of safety practices in the warehouses, issues of decentralised systems for material management; are some of the significant issues among 40 other issues that they have identified via the survey [12]–[15]. Also, another study shows that improving the inventory management system will reduce the cost of materials and logistics, improve

profitability, reduce waste, and reduce damage to stocks [12]. Another study to improve profitability via inventory management has shown that inventory management systems will improve storage optimisation, reduce material purchase costs, reduce or prevent material shortages, and maintain adequate safety stocks [16]. Several studies suggest that manual and traditional warehouse management system has to be replaced with new technologies to benefit the industry, and improper warehouse management practices are common in on-site inventory management, which causes additional costs for the companies [13], [17]. A study shows that procurement fails to place correct orders with correct quantities at the right time, and on-site inventory managers fail to set the proper amount of safety stocks, reorder points, and maximum stock levels due to the usage of traditional inventory management systems in construction supply chains [18].

At the end of the 20th century, Japanese contractors were successful in the market due to their technological superiority. The main reason they became superior in the construction industry was that 0.5% of their revenue was dedicated to Research and Development (R&D) [19]. A survey conducted by Boston Consulting Group with 50 advisory members and industry experts worldwide found that technology transformation is the second most significant transformation the construction industry needs and has a weight of 4.5 out of 5. In recent years it has been noticeable that construction companies invest in IT-related implementations and adaptations. However, most companies still have to fail R&D to check the applicability of new technologies [20]. Countries like Germany, Australia and Singapore improve transparency and smooth processors, reduce corruption and cost reduction, and increase productivity mainly due to implementing new technologies through R&D [20]. A study in Nigeria shows even though inventory and material management are important, especially in terms of project delivery cost and time, in practice, members of the construction industry often fail to use proper techniques [13]. A survey done by Boston Consulting Group in 2016 [21] has found numerous applications of new technologies to optimise, automate, and analyse data worldwide. Some of them are; (a) Skanska has used the wireless sensor to monitor the building progress and reduce the chance of unexpected failures by 50%, improved building management productivity by 20% to 30%, and improved the building's energy performance by 10%, (b) Arup has used data from CCTV footages, traffic data and mobile surveys to get better decision making, (c) Skanska have used Radio Frequency Identification (RFID) for monitor movement of materials and storage which has reduced the total project cost by 10%, (d) Komatsu the heavy machinery manufacture has introduced several automated and semi-automated heavy equipment which uses computer vision with other AI techniques to operate [21].

DL has many applications but not much as ML because the scope and model availability of ML is significantly high compared to DL [6]. However, DL is frequently used in areas like; computer vision, forecasting, image processing, adversarial cases, autonomous vehicles, data mining, Natural Language Processing (NLP), recommender systems, big data analytics, object detection, and visuals generation, dimensionality reduction, classification, pattern identification. The below paragraphs explain some applications in object detection, forecasting, and NLP. Research done by Ruiz-del-Solar et al. [8] suggest several ways that DL can use for robotics in areas like; object detection, objects touching & moving, scene representation and classification, sensing the environment, and spatiotemporal vision. Zhao et al., [9] discuss how DL can be used for computer vision tasks like salient object detection, face detection, and pedestrian detection.

When analysing the research papers on applications of DL for image classification, image processing CNN has a clear dominance, and CNN is frequently used in those areas [22], [23]. Inside a CNN, when an image is input to the network, it breaks into small pieces based on the number of layers, and in a CNN, it is possible to break into every pixel. Convolutional layers are extracted from the input image and find the relationships in input pixels by analysing image properties like the gradient, curves, shapes, texture, colour, edges, contours, strokes, and orientation. When the output of convolutional layers comes to the pooling layer, it reduces the dimensionality of the image outputs of convolutional layers. Finally, connecting layers provide the desired output [6]. CNN is commonly used for image processing, image classification, object detection, and face recognition [10]. CNN neurons are three-dimensional width, height, and depth, where the data is stored as a three-dimensional matrix [10]. CNN has different class variations like LeNet, AlexNet, VGGNet, GoogleNet, ResNet, ZFNet, and ResNet is frequently used for CNN since it can overcome vanishing gradient issues over other conventional CNN classes [6]. Several researchers have highlighted that using CNN for object detection is better than other DL methods, mainly its high performance, ease of developing the model, number of variations, and specific design for visual processing [23]. Also, CNN has promising results on object detection compared to humans [23]. ResNet variation of CNN is commonly used among researchers for object counting, and it can surpass the vanishing gradient issue when there are many layers that other variants cannot address [23], [24]. Also, in a study, researchers suggest that CNN is the best architecture for robotic vision, which performs tasks like object detection and counting [8].

The problem statement of the research is that “Manual and conventional methods of construction supply chain warehouse management incur more costs, take more time, lack real-time data, and struggle to fulfil unanticipated needs. However, Deep

Learning's ability to imagine categorisation and pattern recognition provides compelling solutions. Embracing this technology, like other sectors, offers effective warehouse automation, cost savings, and increased response to changing market needs [25], [26]”.

The identified research gaps are as follows: The following are the research gaps that have been identified: (a) in comparison to other industries, the construction sector lacks digitalisation and adapts slowly to new technologies [26], (b) inadequate research and initiatives focusing on automation in construction inventory management [26]–[28], (c) the prevalence of improper inventory management techniques in the construction industry [13], [29], (d) the disadvantages of relying on manual and traditional inventory management solutions (on-site warehouses) [17], (e) Deep Learning's untapped potential for pattern identification, picture classification, and image modelling in construction inventory management [9], [22], [24], (f) the numerous applications for Deep Learning in construction inventory management [7], [10], [22].

Addressing these research gaps is critical for improving the construction sector's efficiency, productivity, and competitiveness while also harnessing the benefits of modern technologies like Deep Learning.

3. METHODOLOGY

The proposed solution involves implementing an automated system to monitor workers and count bricks in a warehouse using a video feed. For this purpose, three CCTV cameras are strategically placed at the warehouse entrance and inside the warehouse. A review of the literature on deep learning and supply chain research is undertaken to establish the best appropriate methodology, and laboratory experiments are determined as the preferred option due to their effectiveness, time efficiency, and industrial relevance. For model development, a Convolutional Neural Network (CNN) architecture built with Python and Keras is used. Face recognition is performed using 'dlib,' while item counting is achieved using ResNet. A prototype with 100 bricks, three cameras, and a computer is being tested to compare CNN's accuracy to human counting. An image dataset of bricks and human faces was collected, as well as a survey to assess human counting accuracy and expert interviews to validate the research. Accuracy, loss, precision, recall, IoU, MAP, RMSE, and MSE are among the parameters used to train, evaluate, and compare the models. Both the CNN model and human counting are evaluated using the same performance criteria.

3.1. Proposed solution

The solution is to use video feed count and identify the workers who enter the warehouse, who exists, and count the change in the brick block. To get the video feed, three CCTV cameras were set up at the warehouse entrance and inside the warehouse. The first CCTV camera was placed 5 feet from the ground at the entrance facing outwards from the warehouse. The purpose of this CCTV camera was to feed the video feed to the computer to identify the person who entered the warehouse. The second CCTV camera was placed 5 feet from the ground at the entrance facing inwards to the warehouse. The purpose of this CCTV camera was to feed the video feed to the computer to identify the person who exits the warehouse and tally whether the person who enters exits. The third CCTV camera was placed 8 feet above the ground from the ground towards the bricks at an angle where the camera captured the view of 3 sides of the block of bricks. The purpose of this camera was to feed the video feed to the computer to count the number of bricks in the warehouse. A laptop, NVR, or DVR should be in a secure place to capture and store the video feeds (see Figure 1).

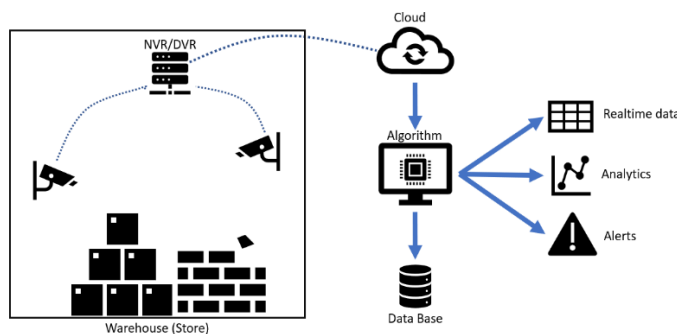


Figure 1: Conceptual framework of the proposed solution

3.2. Feasible methodologies

During the literature, it is found that the most popular approaches for this type of research are laboratory experiments, field experiments, surveys, and case studies. To refine the methodology, 47 research papers on deep learning and supply chain were analysed. The results showed that 64.1% of the research used laboratory experiments as the main approach to the methodology. 22.4% of the sample research papers have used field experiments as methodology. At the same time, surveys and case studies were used 8.9% and 4.5%, respectively, as the methodology. During the analysis, laboratory experiments and field experiments were more concentrated on algorithm building and programming aspects, mainly using primary data, while surveys and case studies concentrated on more theoretical and minor programming aspects using secondary data. In addition to finding the most suitable and most used methodology,

the feasibility and limitations of each method were analysed using the same sample of research papers. From this, it was clear that (a) a laboratory experiment is the feasible method for assess the accuracy of the model considering the time, cost limitations, and dependence on the industry, (b) a survey to assess the counting done by the humans since there are no secondary data nor industry support, and (c) laboratory experiments more flexibility in generalising the findings and conducting [4], [30]–[33]. Then, an image-based survey was conducted to find the accuracy of counting done by humans. Then it was evaluated and compared with the accuracy of the deep learning model.

3.3. Construction of the model

This study selected CNN as the deep learning architecture for the model base on the ease of training and availability of the sample programs. Then python is the programming language, and Jupiter Notebook is the coding platform for model building. This study used Keras 2.1.2 in the model's front end and Theano 1.1.0 in the back end of PyTorch [34]. This study used the 'ageitgey' code as the base code for the face recognition section and the 'gauthsvenkat' code as the base code for the object counting part from each image dataset 90% of the data was used as the training data, and 10% as the testing data.

As for this study, the object detection dataset was classified into 21 classes and selected multiclass classification for the model. The supervised learning method was adopted by labelling every type with the number of bricks and setting the tolerance to 5 since the gap between classes is five bricks. ResNet was selected as the architecture for this study algorithm since conventional CNN architectures face the problem of the vanishing or exploding gradient when the number of layers is increased. The built model has a maximum of 101 layers, four pooling layers, three dropout layers, and four fully connected layers. Flatten layers were used to flatten two-dimensional image data to make input to fully connected layers feasible. The activation function was Rectified Linear Unit (ReLU), and MaxPool was used for pooling layers. Kalahiki [35] proves that ReLU is the best activation function for object counting, among other commonly used activation functions like Linear, Leaky ReLU, and SELU. We used zero-padding to convolutional layers to match the input and output sizes. We used two-dimensional max pooling with size (2,2) at pooling layers. The output size of the first fully connected layer is 640, and the last is 20, which is the number of classes. The cross-entropy function was used as the loss function during the model training and Adam as the optimisation algorithm. The specifications of the computer that was used to develop and train the model had an i7 3.8GHz processor, two Nvidia RTX 3050 VGA, 64GB RAM, 4TB HDD, and 1TB SSD.

3.4. Setting the prototype

As the first step of the solution, a prototype has been set up for testing, and based on the results in future studies; the solution suggested can be fully implemented. As the prototype, it was required to assess the accuracy of detecting bricks via CNN compared to count done by a human. Therefore, 100 bricks, three cameras, and a computer were arranged to test the prototype (Figure 2). Three web cameras had been set up in the arbitrary warehouse's entrance area to get the video feed. Camera 01 was a laptop web camera placed 5 feet above the ground at the arbitrary warehouse's entrance area facing the brick block outwards (Figure 2). The purpose of this web camera was to feed the video feed to the computer to identify the person who entered the warehouse. Camera 02 was placed 5 feet from the ground and attached to the laptop (Figure 2). The purpose of this web camera was to feed the video feed to the computer to identify the person who exits the warehouse and tally whether the person who enters exits. Camera 03 was placed 8 feet above the ground from the ground towards the bricks at an angle where the web camera captured the view of 3 sides of the block of bricks (Figure 2). The purpose of this web camera was to feed the video feed to the computer to count the number of bricks in the block. The same computer that attached the first and second cameras was used to collect video feeds from the cameras.



Figure 2: Camera setup and orientation

As explained above, the cameras will get a video feed and forward it to the model. Once the face is recognised in Camera 01, it triggers the counting algorithm and counts the number of bricks in the warehouse via the video feed from Camera 03. Then once the face is detected in Camera 02, it triggers the counting algorithm again and counts the number of bricks in the warehouse via the video feed from Camera 03. A simple, intermediate program will collect data from algorithms, and based on the data, additional features like warnings, alerts, and information are displayed in the main interface.

3.5. Data collection

For the research, an image dataset of bricks storage with 0 to 100 bricks and images of the faces of 4 humans were used to train the model. Also, 24 responses were collected from the survey, which was conducted to measure the accuracy and the response time of human brick counting. Then, finally, to validate the research significance, we conducted eight expert interviews and analysed the contents of the interviews. The first task was to train the brick counting model using the 12,600-image dataset. The image dataset was generated using web cameras from the artificial warehouses, which were set in the laboratory environment. The whole image dataset was captured from two angles, and the number of bricks was changed from 0 to 100 in intervals of 5 and got 21 classes of images; each class had 600 images. During image set generation, the positions of some bricks were changed to give a more realistic environment to train the model. Also, 120 images (20%) of each class of images were set on the left-right flip to get more variation for the image data set. 6,000 images were collected for face recognition, and 2000 images (33.3%) of the data set for face recognition were generated by four people using web cameras from the artificial warehouses. The rest of the image dataset for face recognition was imported from the existing databases from GitHub, which was generated from eight humans. Each face had 500 images from different angles, and some images were set on left-right inversion to get more variation for the image data set.

Initially, it was decided to collect responses from 30 individuals, but collected only 24 responses. 20 (83.3%) of this sample were from the industry, working as masons or labourers; the other 4 (16.7%) were site supervisors. The average years of experience in the field of the sample are 36.9 years. To give a fair chance to humans, the images of the brick block were shown on a 21-inch television screen, so the individuals could see the bricks clearly and provide accurate and quick answers. The images were presented individually, and asked the interviewee to count the number of bricks in the image and give the answers within 20 seconds. During the answering, the time taken to answer was measured, and the interviewees marked the number of bricks on the provided sheet.

Then, eight online expert interviews were conducted via Zoom. The average years of experience are 16.4 years, and all of the experts are from senior or higher positions in the company. A questionnaire was used as the agenda for the discussion. Then presented a presentation that explained deep learning, the suggested automated system from this study, cost, and technical aspects of the proposed system. The technical elements of those three solutions and the operations, maintenance, and financial investment were explained to experts. After the explanation, we evaluated the willingness and benefits of implementing such a system considering cost,

accuracy, ease of implementation, and possible issues that may arise once this system is introduced to the company.

3.6. Training of object counting and face recognition

The loss, accuracy, precision, recall, IoU, MAP, RMSE, and MSE were calculated from the model itself. Also, during the training, it was found that optimum results are generated between epochs 1 to 100. When the model surpasses that threshold, the model gets deoptimised, and it was decided to stop passing in the neural network because it will be a waste of time and resources [36]. The face recognition model's loss, accuracy, precision, IoU, and MAP were calculated from the model itself. However, since no numbers are involved in this algorithm, it was not possible to calculate any of the values for the RMSE and MSE. During the training, it was found that optimum results are generated between epochs 1 to 30. The data set was trained with 6,000 images.

3.7. Evaluation of model performance

The equation shown in Equation 1 is used to calculate the model's accuracy.

$$Accuracy = \frac{TP}{TP+FP} \quad (1)$$

Where TP denotes True Predictions, and FP denotes False Predictions. The loss, accuracy, precision, recall Intersection over Union (IoU) and Mean Average Precision (MAP). In addition to that, the Root Mean Squared Error (RMSE), and Mean Squared Error (MSE) were calculated for each model from the model itself to get a statistical value to compare [37] with CNN and counting done by humans. The evaluation thresholds of each parameter are explained in Table 1.

Table 1: Checked parameters to evaluate models [23], [35], [37]

Parameter	Output type	Indication	Evaluation Thresholds
Loss	Graph	Should be concentrated closer to 0	-
Accuracy	Graph	It should be concentrated closer to 100% or 1	-
Intersection over Union (IoU)	Graph	It should be concentrated closer to 1	-
Mean Average Precision (MAP)	Ratio/Percentage	Should be closer to 0.01 or 1% but practically it goes around 30% to 50%	Less than 50%
High Confidence False Positives	Ratio/Percentage	Should be closer to 0.01 or 1%	Less than 5%

Root Mean Squared Error (RMSE)	Value	It should be closer to 0 and more significant than MSE but should be closer to MSE	Less than 2
Mean Squared Error (MSE)	Value	It should be closer to 0 and lower than RMSE but should be closer to RMSE	Less than 2

3.8. Evaluation of human counting accuracy

Minitab and Excel have been used to get the graphs, scatter plots, and calculate RMSE and (MSE) of human counting. RMSE is calculated using Equation 2, and MSE is calculated using Equation 3. RMSE and MSE have been used to compare the counting performance of different methods [38]. Calculating the same criteria for both the CNN model and human counting will provide a common platform to evaluate each's performance. RMSE is always a positive number, with "0" denoting the ideal match, which is very rare in real-world applications. Additionally, it is a scale-dependent accuracy metric [39]. Bar graphs are used to identify the human counting accuracy for each class and then compared with CNN counting accuracy for each class. The scatter plot is used to understand how the human counting input has varied,

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}} \quad (2)$$

where,

\hat{x}_i : predicted value

x_i : actual value

n: the size of the data set

MSE is widely used when the prediction values have a wide range and are not limited to a specific scale. It is a common choice for evaluating the accuracy of predictions in various fields, including forecasting and machine learning [40]. MSE provides a measure of the average squared difference between the predicted values and the actual values. It is particularly useful when the magnitude of errors is important in assessing the accuracy of predictions.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (3)$$

where,

\hat{Y}_i : predicted value

Y_i : actual value

n: the size of the data set

4. RESULTS AND DISCUSSION

4.1. Analysing results of object detection by CNN

The 6000-image dataset was split into 90% for training and 10% for testing to find the best possible accuracy, loss, MAP, and High Confidence FP. The maximum accuracy of the training and testing data has reached 99%, and 92% respectively which is a perfect amount (Figure 3). Between the 15th and 100th epochs, accuracy varies between 80% to 90%, which indicates that accuracy is comparatively good (Figure 3) [35]. The accuracy pattern of the training data does not show any characteristics of an overfit [36]. The loss in training data is lower than 0.5 around 80% of the epochs, which is very low than the tolerance amount (Figure 4). When considering the testing data, more than 90% of the epochs show less than one loss value (Figure 4). The loss trend of training data does not offer any overfit characteristics.

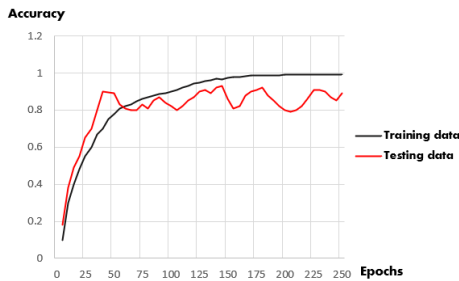


Figure 3: Transition of the accuracy of training and testing data in the data set

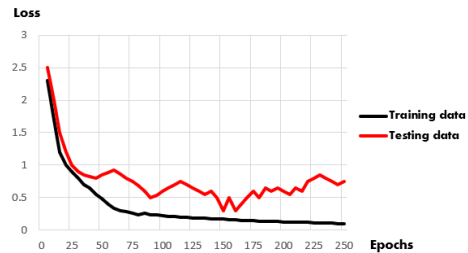


Figure 3: Transition of loss of training and testing data in the data set

This model has reached its best accuracy and performance. Only 0.2% of High Confidence FP indicates only one false positive out of 600 testing images. 47.9% of MAP is lower than the set threshold and 2.1% lower than the selected threshold. Low MAP shows that the model has achieved a censurably good precision. RMSE (0.708) and MSE (0.793) values are very low, and the gap between RMSE and MSE is closer to each other, which is 12.0% of the RMSE. All these data suggest that the model is very accurate and still has no significant number of errors.

4.2. Analysing results of face detection by CNN

For this study, the algorithm was trained using the data set that generated in the laboratory environment, which consists of 6000 images. The original developer had trained the algorithm adopted in this study. Because of that, we checked the loss, mAP, and High Confidence FP only up to 30 epochs. The maximum accuracy of the training data has reached 99%, which is an excellent amount [41]. Also, the maximum

accuracy of the testing data has reached around 96%. Between the 8th and 30th epochs, accuracy increases from 80% to 96%, indicating that accuracy is excellent and that there is no need for further training (Figure 5). The model's accuracy pattern of the training or testing data does not show any characteristics of an overfit. The model indicates very low loss, and around 80% of the epochs have a loss lower than 1 in both the testing and training of the model (Figure 6) [42]. The loss trend of the model does not show any characteristics of overfitting. This model has reached its best accuracy and performance. There is only 0.8% of High Confidence FP. 44.7% of MAP is lower than the set threshold and 5.3% lower than the selected threshold. Low MAP shows that the model has achieved good precision.

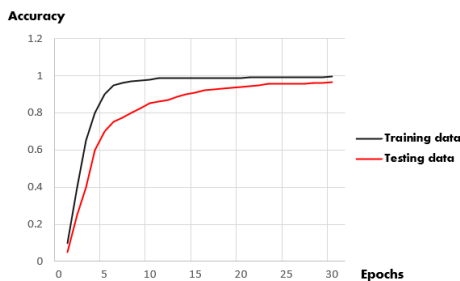


Figure 4: Transition of the accuracy of training and testing of the face recognition model

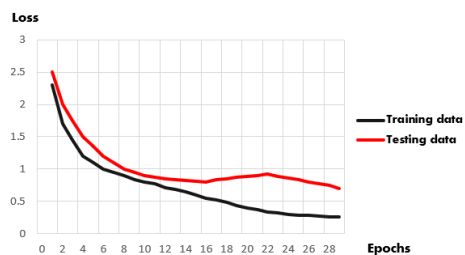


Figure 5: Transition of loss of training and testing of the face recognition model

4.3. Analysing results of counting by a human

During the analysis, a bar graph for the accuracy of the counting done by a human for each class was generated. Humans can guess the number of bricks on 0 bricks and five brick class, but after ten brick class, there is a significant decline in the counting accuracy. From class 20 to class 85, humans cannot count more than 50% with accuracy (Figure 7). However, after class 90, there is an increase in counting accuracy, which stays above 50% accuracy rate, and the mean accuracy of the whole data set is 47.2%.

RMSE (8.39) and MSE (67) are significantly far from the set threshold. These numbers suggest that the counting done by humans is inaccurate and has considerably significant errors. In addition, the RMSE and MSE gap is very large, almost seven times larger than the RMSE value.

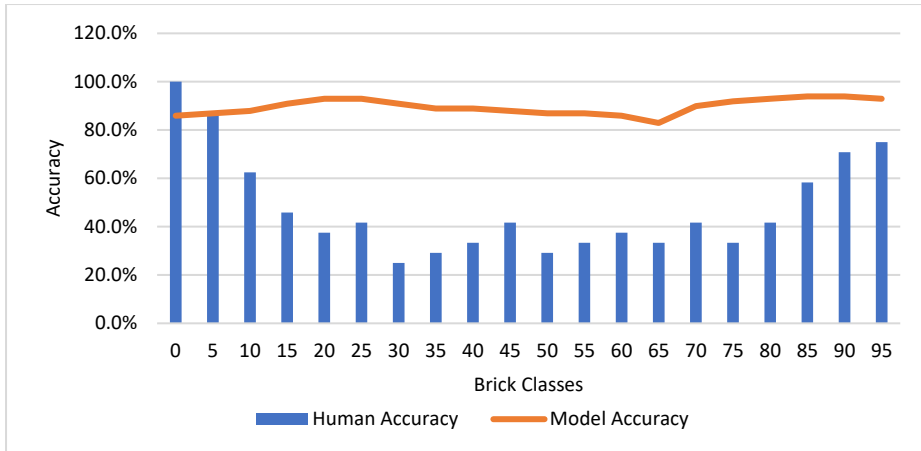


Figure 6: Accuracy of the counting done by a human for each class

4.4. Comparison between CNN and counting by a human

Once the accuracy, loss, and other parameters were calculated, the human counting and CNN model were compared. The CNN model's accuracy is far superior to human counting. The accuracy of the CNN is consistently above 80%, whereas the accuracy of counting done by humans is mostly below 50% (Figure 8). When comparing the RMSE and MSE values of both models, CNN has a better RMSE of 0.708 and MSE of 0.793, whereas human counting has a very high RMSE of 8.39, and MSE of 67 (Table 2). When the ratios of RMSE and MSE between the two models were calculated, it shows that the RMSE of Human counting is 11.85 higher than the CNN model and the MSE of Human counting is 84.49 times higher than the CNN model. All these graphs, data, and parameters show that CNN is better than counting done by a human.

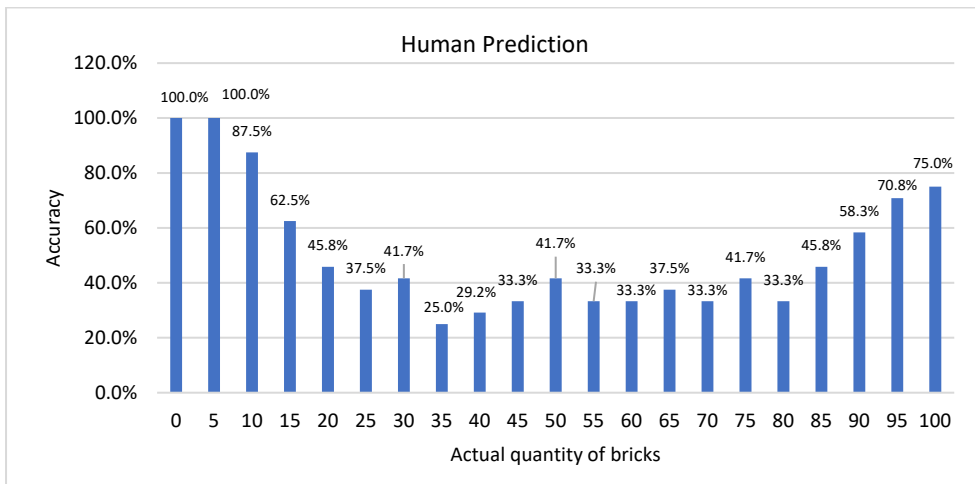


Figure 7: Accuracy comparison of human counting and CNN counting

Table 2: Evaluation metric values of the human counting and CNN counting

Parameter	Human Counting	CNN counting	Human/CNN Counting
RMSE	8.39	0.708	11.85
MSE	67	0.793	84.49

4.5. Counting using CNN

For this study, it was required to evaluate the feasibility and the application of the model. So, two classes were selected where the model's accuracy is low, added 400 newly collected images to each class, and checked the distribution of guesses done by the model [43]. From this, it was possible to get a good idea of the model's accuracy on its weakest points. It was found that class 35 and class 65 have low accuracy compared to the other courses. Then 400 new images with 51 and 53 bricks were added to see whether the model could identify the class to which these images belong.

When we got the results, it was clear that even though these classes' accuracy is low, they are still more than 80% and perform better than humans. It was found that with the new dataset, the model got an accuracy of 90.3% for class 35 (Figure 9). It was found that with the new dataset, the model got an accuracy of 83.25% for class 65 (Figure 10). It is found that even with a little deviation from a class, the accuracy of the brick count does not change. When 400 images of 51 bricks were used, the model identified 59.7% and 28.3% of photos as 50 class and 55 class with 88% accuracy by allocating to 50 and 55 classes (Figure 11). A similar result was given when 400 images of 53 bricks were inputted. The model will identify 33% and 53.7% of photos as 50 class and 55 class, with 86.75% accuracy by allocating to 50 and 55 classes (Figure 12).

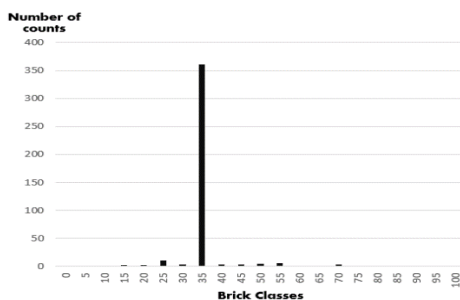


Figure 9: Distribution of the additional 400 images from class 35 which tested with the model

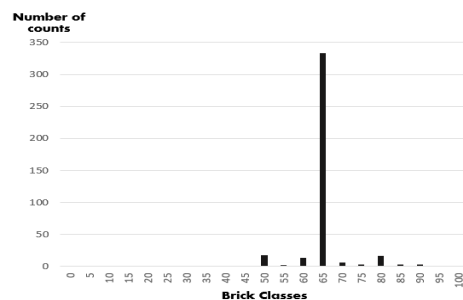


Figure 10: Distribution of the additional 400 images from class 65 which tested with the model

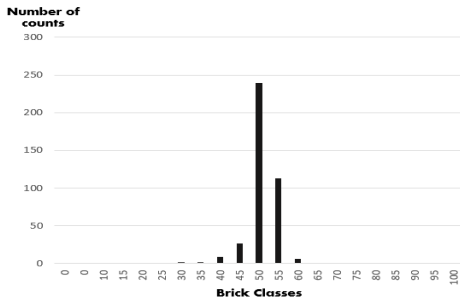


Figure 11: Distribution of the additional 400 images from class 51 which tested with the model

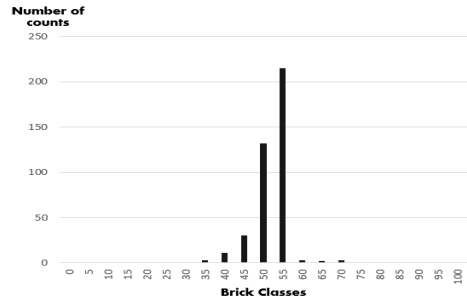


Figure 12: Distribution of the additional 400 images from class 53 which tested with the model

4.6. Outputs of the Model and Validation

The several outputs of the model are shown below. Figure 13 was entirely created by a coauthor for this article, has never been published, and was not based in whole or in part on a previously published figure.

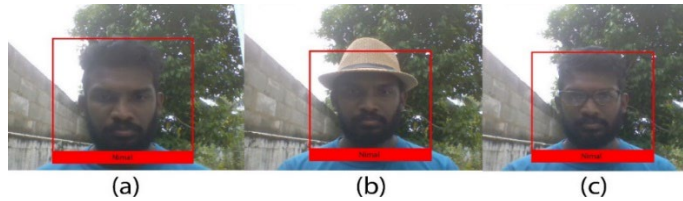


Figure 13: Person identification by the modal with hats, eye wares

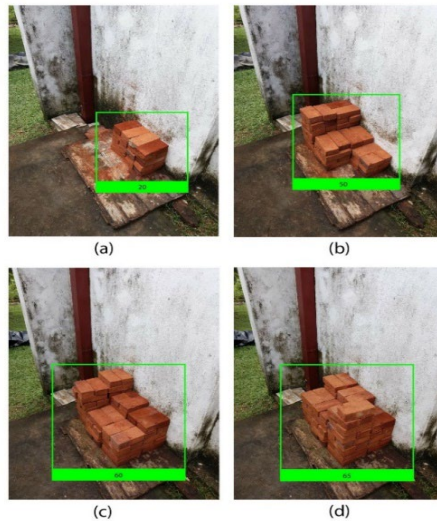


Figure 14: Number of bricks by the object counting modal: a-20, b- 50, c-60, d- 65 bricks

After analysing the content of expert interviews using content analysis methods [44], [45], the main benefits and issues of applying this solution in construction warehouses were identified. According to the experts, labourers must get permission from the gang supervisor, site supervisor, or engineer to get materials from the store. To acquire the materials, labourers must fill out the 'Material Requisition Form' 'Gate Pass' to get materials, which consumes time and reduces efficiency. Labours must wait until the storekeeper counts, inspects, and issues the materials; this also takes time. The system will reduce the data entry, order placing, document maintenance, and bill card maintenance of storekeepers and purchase officers. Also, storekeepers do not have to enter data and take time to generate reports for the HQ; instead, the system will do the reporting. Since the system identifies each person who comes to the store and takes, it is easy to backtrack if wanted materials are misplaced. So, the responsibility for the materials will have to bear by the labourers. This system can be used in off-site warehouses like prefabrication and main warehouses. Due to the availability of real-time data, higher management can track each site's inventory levels. They can direct excess materials to that site if there is an emergency requirement. Since this system can give alerts about order points and stock levels, it is easy to place orders on time accurately. Some companies can market implementing these systems to showcase that they are using advanced technologies. Since the time in the store and airing material is low and the store is always open, materials issues per day can increase. Using the digital data generated by the system will enable the company to adopt new technologies or integrate them into its existing systems.

More than 80% of the experts believe that this system will reduce the time for material acquisition and paperwork and overcome the manual store closures on the work site (Table 3). Also, more than 50% of experts believe this will ease the work of purchase officers and storekeepers, track inventory, identify who is responsible for materials, and use off-site warehouses. There are a total of 14 benefits from this system (Table 3).

Table 3: Benefits highlighted by experts and the response percentage

Benefits	The percentage from the sample
Save time for material acquisition	87.5%
Reduce paperwork	87.5%
Overcome the issues of store operating times	87.5%
Ease the work of storekeepers and purchasing officer	75.0%
Ease of tracking inventory level	75.0%
Appointment of responsible person for materials	75.0%

Ability to use this in prefabrication and main stores	62.5%
Higher management can track inventory levels	50.0%
Ease of placing orders on time and correctly	50.0%
The company can market this	37.5%
Ability to issue materials frequently	37.5%
Ability to use other technologies	37.5%
Ability to collect materials quickly from the store	25.0%
Generation of digital data	25.0%

No more than 50% of the experts have mentioned common issues experts see in implementing the system (Table 4). 50% of the experts have noted that the cost of such a system is very high, employees in the industry do not have good technical knowledge about these technologies, sometimes the materials are stored in different locations due to lack of space to keep them, and the system might face issues addressing this (Table 4). The experts mention seven issues if this system implements in the construction warehouses. All the SME companies and government institutes said they have financial issues implementing this.

Table 4: Issues highlighted by experts and the response percentage

Issues	The percentage from the sample
Financial limitations	50.0%
Lack of technical knowledge of employees in the construction industry	50.0%
Materials are stored in several locations within the site	50.0%
The job security of the storekeepers	37.5%
The store is moving along with the project location moves	37.5%
Issues regarding the CCTV system	25.0%
Resistance from the workers	12.5%

Implementation of this system may benefit the industry in many ways, and that has become a fact due to the feedback from industry experts.

4.7. Implications for Research, Practice, and/or Society

1. Researchers can investigate and evaluate the performance of various deep learning (DL) models for automating construction inventory management in

future studies. Such research could yield useful insights into the most effective ways for this aim.

2. Another possibility is to investigate DL models' capacity to count various sorts of materials at the same time. This would increase the inventory management system's adaptability.
3. The model was trained to count a small number of bricks using high-quality photos during this study. Future research should look into the DL model's capacity to count a high number of material units properly, even when given low-quality images.
4. Additionally, the current study only used one camera for brick counting, restricting its range of view. Using numerous cameras from different angles in future investigations might result in more precise and thorough material counts.
5. Another possible direction for future study is to look at DL's ability to count and identify persons inside a single camera view. This may reduce the requirement for separate cameras to monitor entering and exiting workers.
6. The presence of several workers handling products in the warehouse at the same time is an important concern for real-world applications. In future studies, researchers should evaluate the DL model's and/or automated system's capacity to properly compute and monitor material flow under such dynamic situations.
7. Researchers could compare and contrast how DL is used in inventory management systems across different sectors to widen the spectrum of possibilities. This analysis could lead to new applications and enhancements for this study.
8. Future research can collect and analyse more input and ideas from different layers of the building sector to further confirm this study and uncover practical issues and improvements. This comprises industry experts, front-line employees, workers, suppliers, inventory managers, project managers, and businesses of all sizes and scopes.

Given the enormous number of materials in construction warehouses and the difficulties in obtaining high-quality images, the findings from these areas of research provide numerous opportunities to develop a DL-based automated inventory management system that caters to the construction industry's unique demands.

4.8. Limitations and future research directions

Several limitations and challenges were faced when building the model, evaluating the model, and validating it.

- a) With the time limitations and to stick to the main scope. This study did not analyse many research papers to find the best algorithm for this model; instead, the papers closer to the search results were selected. There is a possibility of analysing more research papers and coming to a more solid conclusion. In the future, an extensive paper analysis will help get more insights into the most suitable model.
- b) For the model's simplicity, the number of classes was reduced, visualising area, the number of activities detected, and separate cameras for each task. With more data, resources, and time, it is possible to develop a more advanced prototype or a solution for warehouse automation in the construction industry.
- c) For this study there was little support from the industry during the model development phase and final validation; because of that, for the validation, we only had six experts to interview and get their opinion on using CNN base computer vision for warehouse automation. If it is possible to evaluate the model frequently and by more experts, it provides more insights, benefits, challenges, and validation of CNN-based computer vision for warehouse automation.

There are possibilities to overcome the above limitations and continue this research for warehouse automation in the future, but that would require more resources, support from the industry, and technical capabilities.

5. CONCLUSIONS

In the literature, it was found that the construction industry faces many issues due to traditional and manual methods of managing their inventories and warehouses. A Deep Learning based automated warehouse management in the construction industry was proposed in this study. During the investigation, it was found that the best algorithm for this task is CNN. After training, it shows significantly better accuracy compared to counting done by a human. Also, the research showed that CNN could be used for face recognition, which can be integrated into the warehouse automation system to help identify how much material was borrowed by each worker. In addition to that, using the data from object counting and face recognition, several additional alerts and information could be provided to the system users. In several stages of the research, industry experts were interviewed and got feedback for the proposed Deep

Learning based automated warehouse management and validated that such a system can replace all the manual processors done in a traditional warehouse and improve the industry. Through the expert interviews, it was found that this system industry can benefit in many ways. According to the analysis, most experts have mentioned that this system and most the experts believe that (a) time for material acquisition will be reduced, (b) reduce paperwork, (c) overcome issues of store operating times, (d) the job of the store keeps and purchasing officers will be eased, (e) give the ability to track real-time data of inventories, (f) assign the responsibility of materials to the person who borrowed it, (g) this system can be used for warehouses in on-site or off-site. Other than these, experts have mentioned seven other benefits of this system. Only half of the experts mentioned that implementation of this system would cause issues like (a) financial difficulties implementing this system, (b) lack of technical knowledge of employees in the construction industry, and (c) issues related to monitoring materials which are stored in several locations within site due to unavailability of space and four other issues that less than half of the experts mentioned.

REFERENCES

- [1] P. Childerkouse, J. Lewis, M. Naim, and D. R. Towill, "Re-engineering a construction supply chain: A material flow control approach," *Supply Chain Manag.*, vol. 8, no. 4, pp. 395–406, 2003, doi: 10.1108/13598540310490143.
- [2] M. N. S. N. Jaffar, A. H. Abdul Tharim, "Factors of conflict in construction industry: A literature review," *Procedia Eng.*, vol. 20, pp. 193–202, 2011, doi: 10.1016/j.proeng.2011.11.156.
- [3] O. C. Y.G. Sandanayake, T.B. Dissanayake, "Construction Supply Chain Resilience in Catastrophic Events," *Risk Manag. Treatise Eng. Pract.*, 2018, doi: 10.5772/intechopen.79963.
- [4] Y. Liu, J. Dong, and L. Shen, "A conceptual development framework for prefabricated construction supply chain management: An integrated overview," *Sustain.*, vol. 12, no. 6, 2020, doi: 10.3390/su12051878.
- [5] H. P. Tserng, R. J. Dzung, Y. C. Lin, and S. T. Lin, "Mobile construction supply chain management using PDA and Bar Codes," *Comput. Civ. Infrastruct. Eng.*, vol. 20, no. 4, pp. 242–264, 2005, doi: 10.1111/j.1467-8667.2005.00391.

- [6] S. Sengupta *et al.*, “A review of deep learning with special emphasis on architectures, applications and recent trends,” *Knowledge-Based Syst.*, vol. 194, p. 105596, 2020, doi: 10.1016/j.knosys.2020.105596.
- [7] T. D. Akinosho *et al.*, “Deep learning in the construction industry: A review of present status and future innovations,” *J. Build. Eng.*, vol. 32, p. 101827, 2020, doi: 10.1016/j.jobbe.2020.101827.
- [8] “A Survey on Deep Learning Methods for Robot Vision.” .
- [9] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, “Object Detection with Deep Learning: A Review,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, 2019, doi: 10.1109/TNNLS.2018.2876865.
- [10] N. Zhu *et al.*, “Deep learning for smart agriculture: Concepts, tools, applications, and opportunities,” *Int. J. Agric. Biol. Eng.*, vol. 11, no. 4, pp. 21–28, 2018, doi: 10.25165/j.ijabe.20181104.4475.
- [11] M. A. Gaviyaa and M. L. Kumar, “Comparison of the Effects of Single and Multi Criteria Model on Inventory Management,” *Int. J. Adv. Res. Eng. Technol.*, vol. 11, no. 6, pp. 950–958, 2020, doi: 10.34218/IJARET.11.6.2020.085.
- [12] H. Patel, D. J. Pitroda, and J. J. Bhavsar, “Analysis of Factor Affecting Material Management and Inventory Management: Survey Of Construction Firms Using Rii Method,” *Int. Conf. “Engineering Issues, Oppor. Challenges Dev.*, no. June, pp. 1–6, 2016.
- [13] D. Ikediashi and G. Udo, “Inventory Management and Construction Project Delivery in Nigeria BT - Collaboration and Integration in Construction, Engineering, Management and Technology,” 2021, pp. 1–7.
- [14] K. Tanskanen, J. Holmstro, J. Elfving, and U. Talvitie, “Vendor-managed-inventory (VMI) in construction,” *Eletronic Libr.*, vol. 34, no. 1, pp. 1–5, 2015.
- [15] V. Weerapura, R. Sugathadasa, M. M. De Silva, I. Nielsen, and A. Thibbotuwawa, “Feasibility of Digital Twins to Manage the Operational Risks in the Production of a Ready-Mix Concrete Plant,” *Buildings*, vol. 13, no. 2, 2023, doi: 10.3390/buildings13020447.
- [16] L. H. Shan, “Maximize Profit by Implementing Inventory Management in Construction Industry,” pp. 0–13, 2020.

- [17] Z. Dakhli and Z. Lafhaj, "Considering Materials Management in Construction: An Exploratory Study," *Logistics*, vol. 2, no. 1, p. 7, 2018, doi: 10.3390/logistics2010007.
- [18] T. Subramani, V. B. Nair, A. David, B. M. Ghouse, and N. S. Kumar, "A Study of Inventory Management System In Construction Industry," *Int. J. Appl. or Innov. Eng. Manag.*, vol. 6, no. 5, pp. 304–311, 2017, [Online]. Available: www.ijaiem.org.
- [19] A. B. Ngowi, E. Pienaar, A. Talukhaba, and J. Mbachu, "The globalisation of the construction industry - A review," *Build. Environ.*, vol. 40, no. 1, pp. 135–141, 2005, doi: 10.1016/j.buildenv.2004.05.008.
- [20] J. Li, D. Greenwood, and M. Kassem, "Blockchain in the built environment and construction industry: A systematic review, conceptual models and practical use cases," *Autom. Constr.*, vol. 102, no. February, pp. 288–307, 2019, doi: 10.1016/j.autcon.2019.02.005.
- [21] Boston Consulting Group, "Shaping the Future of Construction A Breakthrough in Mindset and Technology," 2016. doi: 10.1061/9780784402474.
- [22] A. Khamparia and K. M. Singh, "A systematic review on deep learning architectures and applications," *Expert Syst.*, vol. 36, no. 3, pp. 1–22, 2019, doi: 10.1111/exsy.12400.
- [23] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021.
- [24] H. Zhang and X. Hong, *Recent progresses on object detection: a brief review*, no. June. Multimedia Tools and Applications, 2019.
- [25] P. H. Chen and T. C. Nguyen, "A BIM-WMS integrated decision support tool for supply chain management in construction," *Autom. Constr.*, vol. 98, no. August 2018, pp. 289–301, 2019, doi: 10.1016/j.autcon.2018.11.019.
- [26] M. Gamlath, K. G. A. S. Waidyasekara, and S. Pandithawatta, "Feasibility of Robotic Technology for the Advancement of Construction Industry in Sri Lanka," *MERCon 2020 - 6th Int. Multidiscip. Moratuwa Eng. Res. Conf. Proc.*, pp. 1–11, 2020, doi: 10.1109/MERCon50084.2020.9185384.

- [27] H. Yan, N. Yang, Y. Peng, and Y. Ren, "Data mining in the construction industry: Present status, opportunities, and future trends," *Autom. Constr.*, vol. 119, no. August 2019, p. 103331, 2020, doi: 10.1016/j.autcon.2020.103331.
- [28] T. Bock, "The future of construction automation: Technological disruption and the upcoming ubiquity of robotics," *Autom. Constr.*, vol. 59, pp. 113–121, 2015, doi: 10.1016/j.autcon.2015.07.022.
- [29] H. E. Pang, R. Chandrashekar, and W. H. N. W. Muda, "Forecasting and economic order quantity model for inventory control: A case study at XYZ company," *AIP Conf. Proc.*, vol. 2184, no. December, 2019, doi: 10.1063/1.5136380.
- [30] C. Deck and V. Smith, "Using laboratory experiments in logistics and supply chain research," *J. Bus. Logist.*, vol. 34, no. 1, pp. 6–14, 2013, doi: 10.1111/jbl.12006.
- [31] R. Hebbalaguppe, G. Garg, E. Hassan, H. Ghosh, and A. Verma, "Telecom inventory management via object recognition & localisation on google street view images," *Proc. - 2017 IEEE Winter Conf. Appl. Comput. Vision, WACV 2017*, pp. 725–733, 2017, doi: 10.1109/WACV.2017.86.
- [32] H. Sun, H. V. Burton, and H. Huang, "Machine learning applications for building structural design and performance assessment: State-of-the-art review," *J. Build. Eng.*, vol. 33, p. 101816, 2021, doi: 10.1016/j.jobbe.2020.101816.
- [33] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, no. December 2012, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.
- [34] M. Shyu, S. Chen, and S. S. Iyengar, "a Survey on Deep Learning Techniques," *Strad Res.*, vol. 7, no. 8, 2020, doi: 10.37896/sr7.8/037.
- [35] B. S. Christopher Bradley Kalahiki, "Computer Vision for Inventory Management," 2020, [Online]. Available: <https://digitalcommons.latech.edu/dissertations/545>.
- [36] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019, doi: 10.1109/ACCESS.2019.2912200.

- [37] A. Almalaq and G. Edwards, “A review of deep learning methods applied on load forecasting,” *Proc. - 16th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2017*, vol. 2017-Decem, pp. 511–516, 2017, doi: 10.1109/ICMLA.2017.0-110.
- [38] S. Kumagai, K. Hotta, and T. Kurita, “Mixture of counting CNNs,” *Mach. Vis. Appl.*, vol. 29, no. 7, pp. 1119–1126, 2018, doi: 10.1007/s00138-018-0955-6.
- [39] J. Wu, G. Yang, X. Yang, B. Xu, L. Han, and Y. Zhu, “Automatic counting of in situ rice seedlings from UAV images based on a deep fully convolutional neural network,” *Remote Sens.*, vol. 11, no. 6, 2019, doi: 10.3390/rs11060691.
- [40] A. Hammoudi, K. Moussaceb, C. Belebchouche, and F. Dahmoune, “Comparison of artificial neural network (ANN) and response surface methodology (RSM) prediction in compressive strength of recycled concrete aggregates,” *Constr. Build. Mater.*, vol. 209, pp. 425–436, 2019, doi: 10.1016/j.conbuildmat.2019.03.119.
- [41] H. Jiang and E. Learned-Miller, “Face Detection with the Faster R-CNN,” *Proc. - 12th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2017 - 1st Int. Work. Adapt. Shot Learn. Gesture Underst. Prod. ASL4GUP 2017, Biometrics Wild, Bwild 2017, Heteroge*, pp. 650–657, 2017, doi: 10.1109/FG.2017.82.
- [42] H. Qin, J. Yan, X. Li, and X. Hu, “Joint Training of Cascaded CNN for Face Detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 3456–3465, 2016, doi: 10.1109/CVPR.2016.376.
- [43] J. Sekhon and C. Fleming, “Towards improved testing for deep learning,” *Proc. - 2019 IEEE/ACM 41st Int. Conf. Softw. Eng. New Ideas Emerg. Results, ICSE-NIER 2019*, pp. 85–88, 2019, doi: 10.1109/ICSE-NIER.2019.00030.
- [44] X. Meng, “Lean management in the context of construction supply chains,” *Int. J. Prod. Res.*, vol. 57, no. 11, pp. 3784–3798, 2019, doi: 10.1080/00207543.2019.1566659.
- [45] S. Don and L. Dinesh, “Lean Strategies To Minimize Waste in Sri Lankan Query Industry,” no. May, 2016.