Automation in Construction xxx (xxxx) xxx-xxx

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: http://ees.elsevier.com

Effective inertial sensor quantity and locations on a body for deep learning-based worker's motion recognition

Kinam Kim, Yong K. Cho *

School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA

ARTICLE INFO

Keywords
Construction worker
Motion recognition
Monitoring
Deep learning
Long short-term memory

ABSTRACT

Construction tasks involve various activities composed of one or more body motions. As construction projects are labor-intensive and heavily rely on manual tasks, understanding the ever-changing behavior and activities is essential to manage construction workers effectively regarding their safety and productivity. While several research efforts have shown promising results in automated motion and activity recognition of the workers using motion sensors, there is still a lack of understanding about how motion sensors' numbers and their locations affect the performance of the recognition, which can contribute to improving the recognition performance and reducing the implementation cost. Moreover, further research is necessary to seek the motion recognition model that accurately identifies various motions using motion sensors attached to the workers' bodies. This study proposes a construction worker's motion recognition model using the Long Short-Term Memory (LSTM) network based on an evaluation of the effectiveness of motion sensors' numbers and locations to maximize motion recognition performance. The evaluation is conducted by generating different datasets containing motion sensor data collected from the sensors located on different body parts. Comparing the performance of five machine learning models trained using the datasets, the desired numbers and locations of motion sensors are identified. The quasi-experimental test with multiple subjects is conducted to validate the findings of the evaluation. Based on the findings, the LSTM network for recognizing construction workers' motions is developed. The LSTM network classifies various motions of the workers that can be utilized as primitive elements for monitoring the workers regarding their safety and productivity.

1. Introduction

One essential way to achieve success in construction projects is to monitor and manage construction workers' activities and working conditions. Since construction projects are labor-intensive and heavily rely on manual tasks, monitoring and understanding the ever-changing activities and motions of the workers are necessary to effectively manage the workers for improving safety and productivity. Construction projects, in general, require excessive and repetitive physical activities, which arouses the strong need for understanding the worker's activities and motions for ensuring and improving the safety and productivity of individual workers.

Posture-related safety risks have been one of the major concerns in construction projects that should be addressed. According to the Construction Chart Book from the center for construction research and training (CPWR) [1], the rate of Work-related MusculoSkeletal Disorders (WMSDs) in the construction industry (34.6 per 10,000 Full-Time Equivalent workers, FTEs) in 2015 was 16% higher than the rate for all industries (29.8 per 10,000 FTEs). WMSDs are the work-related injuries of the muscles, tendons, joints, and nerve tissues [1]. These injuries are caused by repetitive movement, high force exertion, vibration, and awkward body posture, which are frequently observed from construction workers [2]. Although several efforts have been made to mitigate the posture-related safety risks of the workers, existing methods such as training, education, and observation by a site manager are not sufficient to manage the risks. Even identifying the risks from multiple workers at an individual level, which should be prior to deploying safety measures, is still challenging. Therefore, the workers' activities and motions have to be individually monitored to identify and mitigate posture-related safety risks in jobsites.

Meanwhile, the construction industry has experienced a lagging improvement or even a decline in productivity while other industries showed a noticeable improvement in productivity [3]. As the construction industry is naturally labor-intensive, labor productivity directly affects construction productivity [4]. Since 1995, labor productivity growth in global construction industries has been only about 1% while it is 2.8% in the overall industry [5]. Although there is notable labor productivity growth in several sections such as multi-family hous-

* Corresponding author.
E-mail addresses: kkim734@gatech.edu (K. Kim); yong.cho@ce.gatech.edu (Y.K. Cho)

https://doi.org/10.1016/j.autcon.2020.103126
Received 16 September 2019; Received in revised form 3 Febraury 2020; Accepted 6 Febraury 2020
Available online xxx
0926-5805/© 2020.

ing and industrial construction in the U.S., the sections considered accounted for less than 10% of the total hours in construction in 2012 [6]. To improve labor productivity in construction projects, it is essential to measure labor productivity at the individual level. However, the current practices rely on manual methods, e.g., sampling through observation. Those methods are not only labor-intensive but also error-prone, and it is challenging to measure the productivity of multiple workers in the individual level using those methods. To facilitate the individual-level measurement of productivity, monitoring individual workers' activities should be preceded. Since construction tasks inherently involve repetitive physical activities, identification of workers' motions can be a useful clue to understanding the activities of the workers.

This study focused on recognizing individual construction workers' motions such as standing, bending, squatting, and twisting. These motions commonly made by the workers can be used as primary elements to understand working conditions, including the safety and productivity of the workers performing construction tasks. For example, bending and squatting motions can be used to recognize unsafe posture from the workers performing heavy material lifting tasks. Likewise, an idle state of workers can be detected by monitoring workers' motions. This information can be used for analyzing work efficiency and productivity.

This study proposes a construction worker's motion recognition model using the Long Short-Term Memory (LSTM) network based on an evaluation of the effectiveness of motion sensor locations to maximize motion recognition performance. Through multiple implementations of machine learning classifiers with different numbers of motion sensors and their locations, the effectiveness is investigated to find the reasonable numbers and the locations of the sensors. A set of wearable motion sensors including 17 Inertial Measurement Units (IMUs) is used to collect the motion dataset from a subject performing material handling activities such as lifting, carrying, and placing. Five machine learning algorithms for classifying motions from the sensor data are implemented to evaluate the effectiveness of the numbers and locations of the motion sensors. By selecting a part or entire dataset based on the locations, different datasets are generated and utilized to train and test motion classification models. The effectiveness of the number and locations of motion sensors is evaluated by comparing the performance of the models. A new dataset is collected from a quasi-experimental test where multiple subjects perform the same material handling activities. This dataset is used for validating the finding of the evaluation. Based on the findings, the Long-Short Term Memory (LSTM) network for recognizing construction workers' motions is developed. The dataset collected from the quasi-experiment test is used to develop the LSTM model. Using the developed motion recognition model, various motions of construction workers that can be utilized as primitive elements for monitoring the workers can be properly recognized.

2. Motion recognition in construction: current state and knowledge gaps

2.1. Motion recognition in construction

Motion recognition, in general, refers to a pattern recognition method of recognizing human motion states using sensors, such as accelerometer, gyroscopes, and cameras [7]. Motions, here, are defined as postures or actions that humans take during any activities. An activity includes one or multiple motions. For example, a material lifting activity includes multiple motions such as standing, bending, and squatting. Motion recognition has been widely utilized to monitor working conditions regarding the safety and productivity of construction workers. While the state-of-the-art motion recognition techniques generally use machine learning algorithms to recognize motion patterns from the sensor data, motion recognition techniques can be categorized into two approaches based on the type of the sensors used: vision-based approaches and motion sensor-based approaches.

Two approaches have a similar system structure: generating a dataset, extracting features, which is often omitted from the deep learning-based model, training and validating, and testing. However, each approach is distinguishable depending on how to generate a dataset and train a model for obtaining the desired result. The current state of those approaches is reviewed in the following section.

2.1.1. Vision-based motion recognition

Vision-based motion recognition techniques utilize one or multiple vision cameras to capture construction workers' motions or activities. 2D or 3D vision cameras can be used. Images captured by the vision cameras are used as a dataset for training and testing machine learning classifiers that recognize motions or activities of construction workers from the images or for extracting a body skeleton model in a digital format from the images. This approach has been utilized in construction projects [8–19]. Convolutional Neural Networks (CNNs) trained using the pre-trained network and monocular images are utilized to recognize three activities of ironworkers, e.g., walking, transporting, and steel bending [15]. Three channel images including a RGB, optical flow, and gray image are used to extract features from the original images. In addition to recognizing motions and activities, there are several efforts to deploy the motion recognition technique for safety purposes. A deep learning network composed of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) is developed to detect unsafe actions of construction workers from 2D videos [10]. An integrated deep learning model recognizes unsafe actions without extensive preprocessing. 3D joint angles are estimated from a single image using a CNN for ergonomic analysis of construction workers [13]. The trained classifiers in this study recognize motions of body parts such as arm, back, and leg. 3D skeleton models are extracted from a 3D camcorder which includes two lenses to detect unsafe actions of construction workers [12,18]. Han and Lee [18] utilizes a similarity measurement technique which is one of the machine learning algorithms to compare actions with the pre-defined template actions. An action of reaching far to a side on the ladder, which can cause a falling accident due to loss of balance, is selected as a test case in this study. The similarity measurement and comparison techniques for recognizing construction workers' actions are implemented in a case study to validate the unsafe action detection system [8]. 3D joint angles are estimated from skeleton models obtained from depth images [12]. The angles are utilized as criteria to identify unsafe behaviors of construction workers. The value ranges of key joint parameters such as the angle of knees and angle of elbows are determined through a series of experiments using a depth camera [17]. Unsafe behaviors are identified by comparing the joint angles, which are measured by a depth camera to the determined ranges.

Vision-based motion recognition techniques have several advantages of system implementation. First of all, data collecting devices such as 2D and 3D cameras do not disturb construction workers' activities. Comparing to the sensor-based motion recognition, there is no need to attach any device to the human body, which may incur discomfort. Also, existing infrastructure, if available, e.g., CCTV, on jobsites can be utilized to collect data. However, it is still challenging to detect motions from the workers when they are occluded by other objects, or outside of the camera's field of views. Moreover, vision-based methods are sensitive to weather conditions and usually require higher computational capacity than sensor-based methods to process data in training and deploying steps.

2.1.2. Motion sensor-based motion recognition

Motion sensor-based motion recognition techniques use various types of sensors such as accelerometer, gyroscope, and pressure sensor. These sensors are utilized alone or embedded in one device. Recently, Inertial Measurement Units (IMUs) have been widely used because three types of sensors — accelerometer, gyroscope, and magnetometer — are combined in one sensor. By attaching these sensors to subjects' bodies (i.e., construction workers), sensor signals containing information about their movements are collected. Here, one or multiple IMU sensors are utilized to collect the data. The collected data is classified into specific motions or activities. Similar to the typical model development process of the vision-based motion recognition, the sensor data is used for training and testing machine learning classifiers. To enable the classifiers to learn distinguishable characteristics of the data, feature extraction is conducted in a data pre-processing step [20–29]. The feature extraction has been done in general classification tasks using machine learning algorithms to reduce the dimensionality of the data because raw data in high dimensionality contain redundant data which is not necessary for the classification. By selecting and/or integrating variables of the data, the features describing the data efficiently are extracted.

Machine learning classifiers trained using motion sensor data are utilized for various purposes in construction projects. First, unsafe actions of construction workers are detected [20,24,29]. Insole pressure sensors are used to detect construction workers' loss of balance which can cause fall accidents [24]. Five types of machine learning algorithms — Artificial Neural Network (ANN), Decision Tree (DT), Random Forest (RF), k-Nearest Neighbor (k-NN), and Support Vector Machine (SVM) — are implemented in this study. Supervised motion tensor decomposition enables efficient data processing for implementing a SVM classifier [20]. Binary classification on near-miss falls is implemented using one-class SVM with motion data from the wearable IMU sensors [29]. Second, motion and activity recognition is utilized to calculate the productive time of construction workers for understanding productivity [22,23]. Productive time is calculated by classifying activities of the workers using motion sensor data. Lastly, the identification of motions and activities of construction workers is conducted using machine learning algorithms [20,21,26,28,30,31]. [21] uses a smartphone, which embeds IMU inside, attached to a worker's arm to collect motion sensor data. Five types of machine learning algorithms — ANN, DT, k-NN, Logistic Regression, and SVM — are implemented in this study. [28] utilizes a wristband-type accelerometer to recognize activities of masonry work from workers. Different window sizes for data pre-processing are implemented with four machine learning algorithms — k-NN, ANN, DT, and SVM — to investigate their impacts on classification performance. A SVM classifier is developed to classify mason workers into an expert and inexpert group using a set of 17 wearable IMUs [30]. Motion patterns of workers with different levels of expertise are investigated in this study. A convolutional LSTM model recognizes 8 motions of workers using 5 IMUs located on a head, chest, upper arm, right thigh, and right calf [31]. In this study, IMU data in a certain time period are segmented into input data and convolution layers are adopted to the data. Two LSTM layers followed by the convolution layers are used to learn sequential patterns of the data.

Also, motion sensor data are utilized to recognize motions and activities of construction workers without machine learning algorithms [27,32–36]. A network system of wearable IMU sensors enables recognition of construction workers' motions [32]. Motions that can cause musculoskeletal disorders are monitored using two wearable IMU sensors attached to a hardhat and back of safety vest [36]. Tilt angles of the upper body are calculated using IMUs to detect unsafe postures of workers in this study. Fall risk assessment is conducted using IMUs attached to the ankle of construction workers [27,33–35]. Gait stability, which can be an indicator of fall risks, is evaluated using the IMUs.

Motion sensor data contain information about movements of construction workers carrying data collection devices that can be used directly for recognizing motions and activities. Compared to the image data, motion sensor data processing requires smaller computational capacity in data pre-processing, classification model training, and implementation steps. Also, the occlusions made by obstacles do not disturb data collection, which is one of the main disadvantages of vision-based techniques (i.e., line-of-sight issue). Moreover, the impact of lighting conditions on the motion sensor-based system is not significant while the performance of the vision-based system can be affected by the lighting conditions. Other impractical issues, that the vision-based system has, include the fixed camera's limited field of view, disclosing worker's privacy at the unnecessary level, and camera system maintenance (e.g., power supply, cable management, relocation of cameras as construction progresses). In this study, thus, the motion sensor-based approach is selected based on the practical advantages over the vision-based approach.

2.2. Knowledge gaps

Although there have been several efforts to recognize construction workers' motions and activities using motion sensors, the use of the motion sensors is not based on empirical knowledge. To be specific, the locations and the numbers of motion sensors are selected where movements are reflected sufficiently and discomfort from the sensors can be minimized. However, this selection is made without information about how different locations and numbers of sensors can affect motion recognition performance. Moreover, current efforts have low potential to be deployed in different tasks of construction projects because the classifiers trained for recognizing the specific activities are only applicable to the particular task. A limited number of target activities also hinders the deployment of the developed models in complicated and congested tasks that involve various motions and activities. To address these issues, this study proposes an evaluation of the effectiveness of different locations and numbers of motion sensors on motion recognition performance and a deep-learning network that recognizes various motions of construction workers that can be utilized as primitive elements for monitoring the workers.

3. Methodology

This study is composed of three steps; [1] dataset generation and evaluation of the effectiveness of the numbers and locations of motion sensors on motion recognition performance, [2] quasi-experimental test for validation of the findings, and [3] development of LSTM model for recognizing construction workers' motions. Fig. 1 illustrates a framework of the proposed study.

3.1. Evaluation of the effectiveness of the numbers and locations of motion sensors

The evaluation is conducted through the following steps: dataset generation, machine learning algorithm implementation, and identification of the numbers and locations of motion sensors showing the best motion recognition performance.

3.1.1. Dataset generation

1) Target motion classes

The following motions observed from typical construction tasks (e.g., material handling tasks) are selected as target motion classes: standing, bending, squatting, walking, twisting, working overhead, kneeling, and using stairs. Three motions — bending, squatting, and kneeling — are divided into three motions, such as bending-up, bending, and bending-down, to reduce the loss of information caused by the transition of motions, thus 14 motions in total. For example, bending-up and bending-down are transitioning motions from the bending mo -

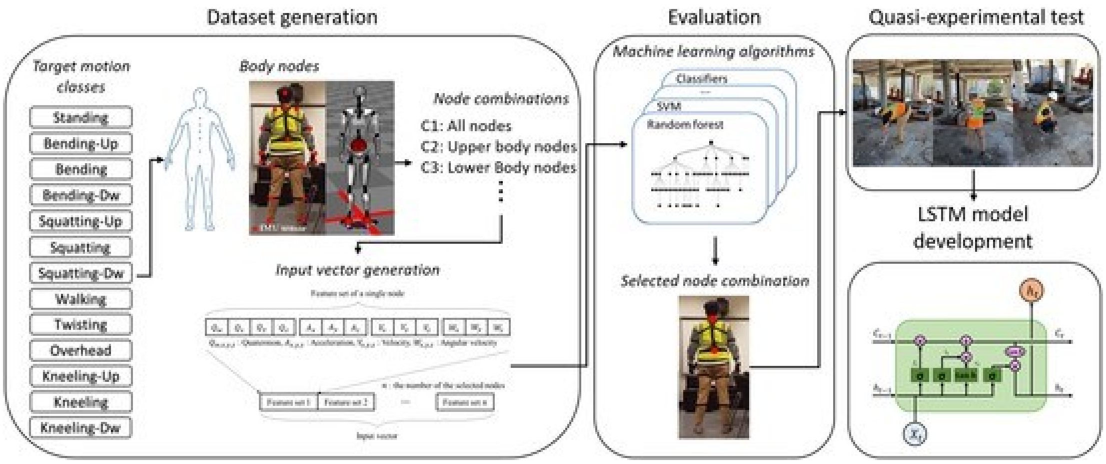

Fig. 1. A framework of the proposed study.

tion to other motions and vice versa, respectively.

2) Body node definition

Twenty-one body joints or body parts throughout the entire body shown in Fig. 2(a) are designated as body nodes for the evaluation. The wearable sensor set of 17 IMUs shown in Fig. 2(b) is utilized to collect the motion sensor data from the nodes. Four nodes located on the spine where sensors do not exist are generated by interpolating between the neck and hip nodes.

3) Node combinations

As shown in Tables 1, 32 node combinations are defined by selecting different numbers of nodes located in different positions. Each combination forms a unique dataset in different dimensions based on the number of the nodes used. By utilizing different nodes for generating datasets, different impacts of the numbers and locations of the sensors are reflected in the datasets. The evaluation includes 32 implementations of machine learning algorithms with 32 node combinations to investigate the performance of motion recognition for evaluating the effectiveness of different numbers and locations of motion sensors on the recognition performance.

4) Input vector formation

Based on the nodes used in each combination, input vectors are defined separately. As shown in Fig. 3, each node with a single IMU generates a feature set composed of 13 values: quaternion (4 values), acceleration (3 values), velocity (3 values), and angular velocity (3 values). These values are calculated using the data from three types of sensors such as accelerometer, gyroscope, and magnetometer embedded in each IMU. Feature sets from multiple nodes in each node combination are concatenated to form an input vector. The length of the input vectors varies depending on the number of nodes in the node combinations. The generated input vectors are used as training and testing data for machine learning classifiers. Each data contains a discrete and time-independent motion state.

3.1.2. Machine learning algorithm implementation

Once the datasets generated from 32 different node combinations are created, five machine learning algorithms — logistic regression, k-nearest neighbor, multilayer perceptron, random forest, and support vector machine — are implemented. These algorithms are supervised machine learning classification algorithms that categorize data from the prior information, i.e., the labeled data [37]. Different types of machine learning classification algorithms that have been widely utilized for motion and activity recognition [21,24,28,29] are deployed to in -

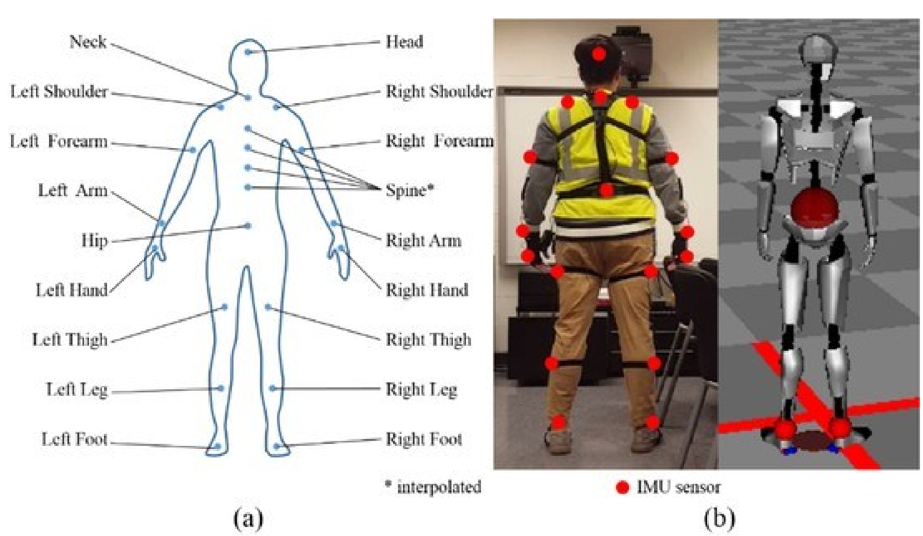

Fig. 2. (a) Defined body nodes; (b) wearable IMUs' locations and their visualization.

Table 1
Node combinations used in the evaluation.

Combination	Selected nodes (the number of nodes)	Combination	Selected nodes (the number of nodes)
1	All nodes (21)	17	Right foot (1)
2	Upper body (15)	18	Left thigh (1)
3	Lower body (7)	19	Left leg (1)
4	Core nodes* (7)	20	Left foot (1)
5	Hip and head (2)	21	Right shoulder (1)
6	Hip and neck (2)	22	Right arm (1)
7	Hip and spine (5)	23	Right forearm (1)
8	Head and neck (2)	24	Right hand (1)
9	Head and spine (5)	25	Left shoulder (1)
10	Neck and spine (5)	26	Left arm (1)
11	Hip (1)	27	Left forearm (1)
12	Head (1)	28	Left hand (1)
13	Neck (1)	29	Spine 3 – close to neck (1)
14	Spine (4)	30	Spine 2 (1)
15	Right thigh (1)	31	Spine 1 (1)
16	Right leg (1)	32	Spine 0 – close to hip (1)

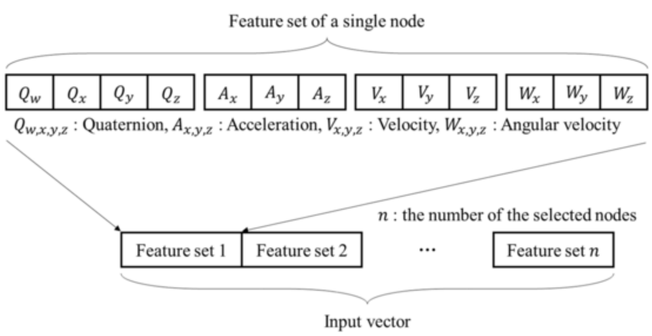

Fig. 3. Input vector formation.

vestigate common findings regardless of the types of the classification algorithms utilized.

Logistic regression is a linear model that calculates the class membership probability for one of two categories in the data [38]. This model is fit by Maximum-likelihood estimation that estimates the coefficients of the model that minimize the error of the calculated probabilities to the one in the data. Logistic regression can handle non-linearity because an activation function is used. In addition, the output of the logistic regression model is interpretable because it is calculated as a probability. K-nearest neighbor is a non-parametric classification method based on the k number data point in the feature space [28,39]. This method assigns an unseen data to a class that has the largest number of data among its k-nearest data point. Euclidean distance in the feature space is the metric to determine the distance between the data points. Hence, the k-nearest neighbor method does not require a training process. Also, this method has only one hyper-parameter, i.e., k (the number of data points to be considered). It enables easy implementation of the method. Multilayer perceptron is a neural network classification model that describes the problems in a network of directed graphs, whose nodes are represented as artificial neurons and the weighted directed edges in the graphs are connections between the neurons [39]. The weights and bias of the network are computed by the backpropagation technique that iteratively updates the weights and bias based on the error rate obtained in the previous iteration. By using a non-linear activation function, multilayer perceptron model can be applied to complicated problems. Moreover, this method is robust to irrelevant input and noise [37]. Random forest is an ensemble method for classification that consists of a combination of decision trees [24,40]. The classification performance of a single decision tree classifier is improved by integrating the bootstrap aggregating (bagging) method and randomization in the selection [40]. This method is fast, scalable, and robust to noise [37]. Support vector machine classifier is a model that classifies data based on the selection of the hyperplane that maximizes the margin between the hyperplane and data points [11]. While the support vector machine is a linear classifier, non-linearity can be handled by using kernels that project the data from the original feature space to a high dimensional space using non-linear kernel function.

A single implementation of the five abovementioned machine learning classifiers using the dataset from the particular node combination is repeated 32 times to simulate all cases, i.e., node combinations. 10-folds cross-validation is used in each implementation to find hyper-parameters of the classifiers and evaluate the classifiers. 10-folds cross-validation is a model validation technique in which the training data is randomly divided into 10 equal parts (folds). Each fold is used as the testing dataset, and remaining folds are used as the training dataset. By doing 10-folds cross-validation, the best hyper-parameters can be properly selected, and the classifiers can be evaluated as generalized classifiers that show the desired classification performance with unseen data.

3.1.3. Identification of the desired numbers and locations of the motion sensors

The desired numbers and locations of motion sensors for construction workers' motion recognition are identified by comparing the classification performance of all classifiers implemented using 32 different datasets. This selection is based on an assumption that the classifier trained using the reduced number of the motion sensors located on particular positions shows comparable classification performance to the classifier trained using all sensors located throughout the entire body. The selected number and location of motion sensors are directly utilized to develop an LSTM motion recognition model for construction workers.

3.2. Quasi-experimental test

The findings of the evaluation are validated through the quasi-experimental test. For the evaluation, the dataset collected from a single subject performing the defined target motions in a lab-environment is used. Since motion patterns vary by individual workers, the dataset collected from a single subject may be biased. Therefore, the new data collected from three subjects involved in the quasi-experimental test is utilized to validate the findings of the evaluation. The subjects are advised to perform typical material handling tasks such as lifting, carrying, and placing in an environment similar to an actual jobsite. While the subject in the evaluation is advised to make predefined motions in which their distribution is balanced, the subjects in the quasi-experimental test are advised to perform typical material handling tasks, such as lifting, carrying, and placing, without considering the balance so that the collected dataset can reflect the characteristics of the tasks. Meanwhile, the raw data of IMUs including acceleration, angular velocity, and magnetic field is used to form input vectors in the quasi-experimental test to minimize the effects of pre-processing the calculation of quaternion and velocity in the evaluation.

3.3. Development of the LSTM model for recognizing construction workers' motions

The five implemented machine learning classifiers categorize each input vector to a class independently. This means that every single moment is classified into a single motion individually. However, the worker's motions have to be interpreted as a result of a sequence of motions. A particular motion can be taken after a sort of motions. For example, bending motion is a result of a sequence of stooping movements from standing or walking motions. In other words, the motions are significantly correlated with the previous motions. This temporal characteristic is considered by extracting features from segmented data using windows in the existing studies [20–29]. However, it is not an effective way to reflect the order of motions in a sequence. Moreover, the selection of the types of features significantly affects the performance of the classifier.

To recognize the motions considering a sequence of previous motions, the LSTM model is developed to recognize construction workers' motions considering a sequence of motions. LSTM is a recurrent neural network designed to learn sequential information using memory cells that stores and outputs information facilitating the learning of temporal relationships on long time scales [41]. LSTM uses the concept of gating that is a mechanism based on pointwise multiplication operations and activation functions, which allow the information to be selectively added to the cell or removed from the cell. Fig. 4 illustrates the basic structure of the LSTM network.

The information flow in the LSTM cell is as follows. First, input values x_i and hidden state from the previous state h_{t-1} pass through the forget gate f_t. The output of the gate is the value between 0 and 1 that represent complete removal of the information and complete retention of the information, respectively. Next, the input gate takes the two values x_i and h_{t-1} to consider new information to be stored in the new cell state C_t. Meanwhile, the values pass through the input modulation gate $\widetilde{C}_t$ with a hyperbolic tangent activation function so that the output value ranges between −1 and 1 that reflects the amount of the information to be forgotten. Subsequently, the old cell state C_{t-1} is updated into the new cell state C_t by multiplying the old cell state and the output of the forget gate, and then by adding the multiplication of the output of the input gate and input modulation gate. After that, the output gate takes input values and the old hidden state, x_i and h_{t-1}, using a sigmoid activation function to decide the parts of the cell state that will be the output. Lastly, the cell state C_t passes through a hyperbolic tangent function, and it is multiplied by the output of the output gate to calculate the new hidden state h_t. Using these gates, the cell state is updated. The equations of the gates and states are as follows:

$$\begin{cases} i_t = \sigma\left(W_{xi}x_t + V_{hi}h_{t-1} + b_i\right) \\ f_t = \sigma\left(W_{xf}x_t + V_{hf}h_{t-1} + b_f\right) \\ o_t = \sigma\left(W_{xo}x_t + V_{ho}h_{t-1} + b_o\right) \\ \widetilde{C}_t = \tanh\left(W_{xc}x_t + V_{hc}h_{t-1} + b_c\right) \\ C_t = f_t \bigotimes C_{t-1} + i_t \bigotimes \widetilde{C}_t \\ h_t = o_t \bigotimes \tanh\left(C_t\right) \end{cases} \tag{1}$$

where σ is the sigmoid function defined as $\sigma(x) = (1 + e^{-x})^{-1}$, $i_t, f_t, o_t, \widetilde{C}, C_t$, and h_t are the outputs of the input gate, forget gate, output gate, input modulation gate, cell state, and hidden state at time t, respectively, $\otimes$ is a pointwise multiplication operator, $W_{xi}, W_{xf}, W_{xo}, W_{xc}, V_{hi}, V_{hf}, V_{ho}, V_{hc}$ are the coefficient matrix, b_i, b_f, b_o, b_c are bias vectors. The coefficient matrix and bias vectors are learnable parameters. By updating these parameters, the model learns the amount of information that passes through the LSTM cell.

In the proposed study, the two-stacked LSTM network is developed in which two LSTM cells are connected to make the network deeper. Fig. 5 illustrates the structure of the developed LSTM network. In the network, the sequence of input vectors that are from the same dataset used in the quasi-experimental test is fed into a fully connected layer followed by a Rectified Linear Unit (ReLU) layer. This technique is im-

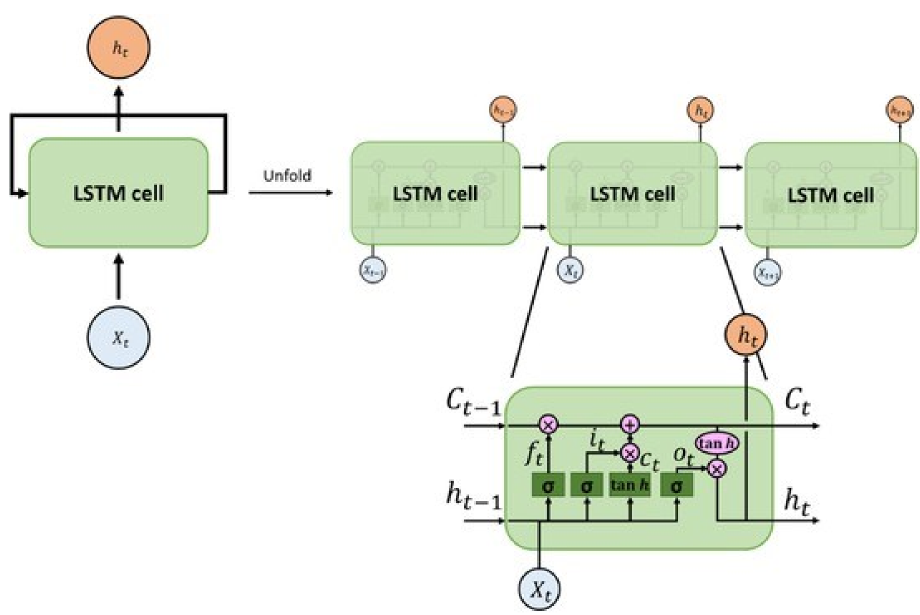

Fig. 4. A structure of the LSTM network.

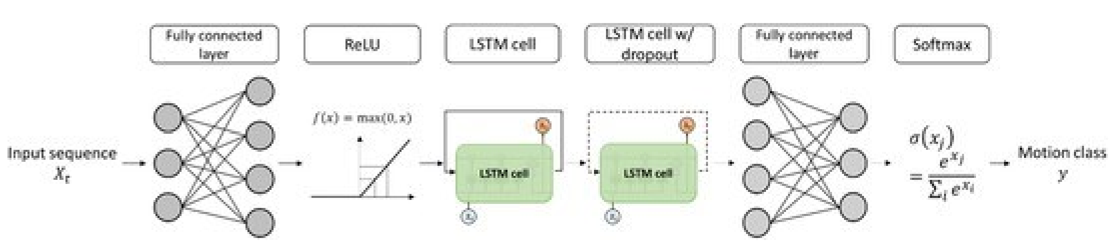

Fig. 5. The structure of the proposed LSTM network.

plemented to improve the network's performance [42]. ReLU is one of the activation functions that have been widely used because it outperforms the sigmoid function. Subsequently, two layers of LSTM cells follow the ReLU layer. Dropout technique is implemented in the second LSTM cells to regularize the network, which allows the network to avoid overfitting and improve performance. In the case of LSTM, the dropout technique excludes the recurrent components such as input, output, and hidden state from the update process probabilistically. At the second LSTM cell layer, the last output of the sequence of the input vectors is only fed into the fully connected layer because the model recognizes the motion at the end of the sequence of motions. Finally, the output of the fully connected layer is fed into the softmax layer to convert class scores into probabilities so that the motion with the highest probability can be identified.

4. Implementation and result

4.1. Evaluation of the effectiveness of the numbers and locations of motion sensors

4.1.1. Data acquisition

A wearable set of 17 IMUs, a commercial product named Perception Neuron, shown in Fig. 2(b) was used to collect the data from a subject performing the defined 14 target motions with a 28-lb concrete block (Fig. 4). In the lab environment, the subject performed material handling tasks including lifting, carrying, and placing with random order of motions. Fig. 6 shows examples of the target motions. Meanwhile, the subject's motions were simultaneously videotaped to be used as the ground truth for labeling the data. In total, 18,350 data points were collected. Each data point has 273 values; each IMU has 13 values and 21 IMUs on 21 nodes. To label the dataset, the authors matched timestamps of the dataset with the recorded videos manually.

4.1.2. Machine learning algorithm implementation

Five machine learning algorithms were implemented using Scikit-learn, which is a machine learning library for the Python programming language. The dataset collected was split into training and testing data that occupy 60% and 40% of the entire dataset, respectively. The dataset was randomly shuffled to minimize the effect of the order in the dataset. Each machine learning algorithm was implemented 32 times to simulate all cases of 32 node combinations. Each implementation included 10-folds cross-validation to find the best hyperparameters. As a result, the overall accuracy of five machine learning classifiers was derived as shown in Fig. 7.

4.1.3. Identification of the desired numbers and locations of the motion sensor

Among five classifiers, a random forest classifier showed the best classification performance in most cases. Table 2 shows the accuracy of the random forest classifier for all node combinations. Based on the assumption that the classifier trained using the reduced number of the motion sensors located on particular positions shows comparable classification performance to the one using all sensors located throughout the entire body, the two classifiers trained using two combinations, combination 5 and combination 6, showed similar classification accuracy. To be specific, the accuracy of the classifier trained using combination 1 that utilizes all nodes is the highest accuracy, 79.83%. Compared to the highest accuracy, the classifiers from node combinations 5 and 6 showed 76.17% and 75.44%, respectively. They are only 3.66% and 4.39% lower than the highest accuracy although they use two nodes, which is only 10% of the total number of nodes of the highest accuracy case.

These two cases have a common characteristic; two nodes are located on the position at a certain distance. This indicates that the data from each node captures the different movement of body parts so that the data can represent different motions effectively and properly. Node combination 8 can be a counterexample. While the number of nodes used in node combination 8 was the same as node combinations 5 and 6, its accuracy was much lower than those of node combinations 5 and 6. It is because node combination 8 includes two nodes located on positions close to each other, which was not capable of representing the motion effectively. This tendency was observed from the results of other machine learning classifiers except the logistic regression classi-

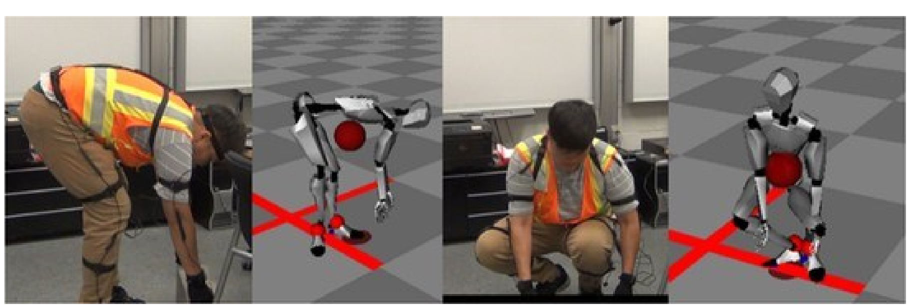

Fig. 6. Examples of the target motions; bending and squatting.

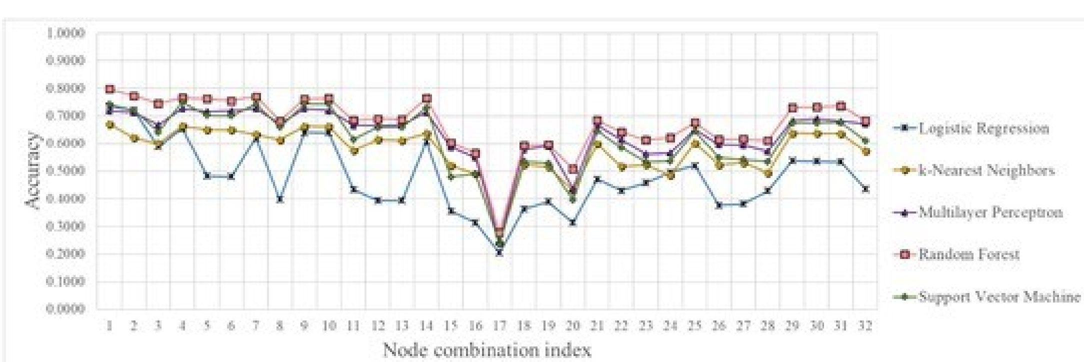

Fig. 7. The accuracy of five machine learning classifiers.

Table 2
Accuracy of Random Forest classifier.

Combi.	Selected nodes (the number of nodes)	Accuracy	Combi.	Selected nodes (the number of nodes)	Accuracy
1	**All nodes (21)**	**0.7983**	17	Right foot (1)	0.2791
2	Upper body (15)	0.7729	18	Left thigh (1)	0.5941
3	Lower body (7)	0.7440	19	Left leg (1)	0.5973
4	Core nodes (7)	0.7658	20	Left foot (1)	0.5090
5	**Hip and head (2)**	**0.7617**	21	Right shoulder (1)	0.6844
6	**Hip and neck (2)**	**0.7544**	22	Right arm (1)	0.6431
7	Hip and spine (5)	0.7713	23	Right forearm (1)	0.6134
8	Head and neck (2)	0.6811	24	Right hand (1)	0.6222
9	Head and spine (5)	0.7606	25	Left shoulder (1)	0.6762
10	Neck and spine (5)	0.7635	26	Left arm (1)	0.6158
11	Hip (1)	0.6849	27	Left forearm (1)	0.6171
12	Head (1)	0.6893	28	Left hand (1)	0.6110
13	Neck (1)	0.6871	29	Spine 3 (1)	0.7298
14	Spine (4)	0.7639	30	Spine 2 (1)	0.7310
15	Right thigh (1)	0.6032	31	Spine 1 (1)	0.7365
16	Right leg (1)	0.5658	32	Spine 0 (1)	0.6833

fier. In the cases of node combinations including one node, the accuracy was significantly lower than the one from node combinations including two nodes. Therefore, the use of two motion sensors located in a certain distance is expected to show similar motion recognition performance to the use of 17 motion sensors located throughout the entire body.

4.2. Quasi-experiment test

Based on the findings of the evaluation, the quasi-experimental test with three subjects was conducted in the environment similar to a construction site. Hip and neck from node combination 6 were selected as the sensor mounting position because their locations are easy to mount sensors. To collect the data from the subjects, data collection devices developed by Robotics and Intelligent Construction Automation Laboratory (RICAL) group at Georgia Institute of Technology were utilized. The devices were carried by the subjects wearing the safety vests with two pockets on the neck and hip as shown in Fig. 8. The devices have a wireless communication module for Wi-Fi and Bluetooth communications, a processing unit, and IMU so that they can automatically upload and store the IMU data to a cloud server. Three subjects were asked to perform material handling tasks including lifting, carrying, and placing with a 28-lb concrete block same as the task performed in the evaluation.

A dataset containing 32,396 data points was collected from the subjects. The data from two devices were concatenated to form input vectors and normalized to have a unit norm. Thus, each data point has 18 values; each data point from one device contains 9 values (3-axis acceleration, 3-axis gyroscope, and 3-axis magnetic field). Then, the dataset was also shuffled and split into training data and testing data that occupy 60% and 40% of the entire dataset, respectively. The random forest classifier that showed the highest accuracy in the evaluation was implemented using the Scikit-learn library. 10-folds cross-validation was utilized to find the hyper-parameters.

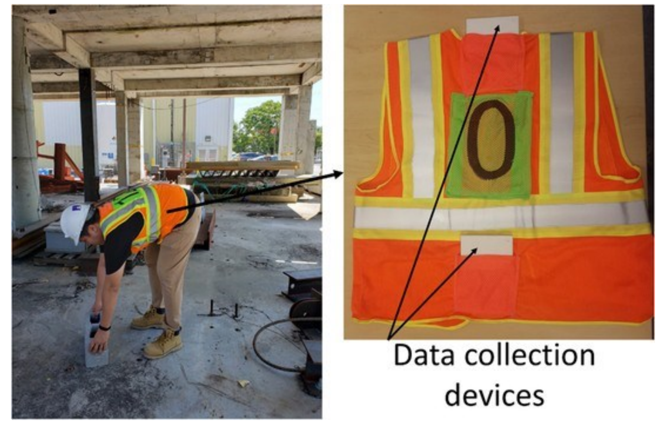

Fig. 8. The quasi-experimental test environment and data collection method.

As a result of the classifier implementation, the confusion matrix was derived as shown in Fig. 9. The heat-map type of confusion matrix shows the results of the classification and the number of each case. The vertical axis indicates the true labels, and the horizontal axis indicates the predicted label. The diagonal values are the number of the instances classified correctly while the off-diagonal values are the number of the instances classified incorrectly. The accuracy, weighted precision, and weighted recall are 82.39%, 83.02%, and 83.10%, respectively. The weighted precision and recall are calculated considering the number of data in each class. 'Using stairs' motions were not observed in the test. Comparing this accuracy to the accuracy of the classifier in the evaluation, the classifier trained using the dataset of the quasi-experimental test shows similar classification performance.

4.3. Development of the LSTM model for recognizing construction workers' motions

The LSTM motion recognition model was developed using Tensorflow, which is an artificial intelligence library using data flow graphs to build models. A computer equipped with Intel® Core(TM) i7-8650U CPU, Intel® UHD Graphics 620, and 16 GB RAM was used to implement the model. The model utilized the same dataset collected from the quasi-experimental test. After two data from two data collection devices were concatenated and normalized, 40 input vectors were segmented into a sequence as an input of the LSTM model. Since the 30 data points were collected in each second, one sequence contains the data collected every 1.3 s. The overlap ratio of the grouping was 95%, which means 95% of the data are shared between adjacent sequences. As a result of the sequence segmentation, 16,178 data points (i.e., 16,178 sequences) were obtained. The shape of the generated sequential dataset was 16,178 by 18 by 40. Once the dataset was modified in the format of the sequential data, the dataset was shuffled and split into training data and test data that occupy 70% and 30% of the total dataset, respectively. Then, the training data was split again into training data and validation data; the training data was used in the actual learning process and the validation data was used in the fine-tuning process. The test data was used to evaluate the performance of the model.

The tuned hyper-parameters of the model are as shown in Table 3. The parameters were tuned by adjusting the values while observing the optimization loss and accuracy of the train and validation sets. As a result, the losses and accuracy over iterations with the tuned hyper-parameters were recorded as shown in Figs. 10 and 11. The losses over iteration graphs showed that the losses were converged enough after 350 epochs, and the difference between train loss and validation loss was small enough, which means the model was well-trained without over-

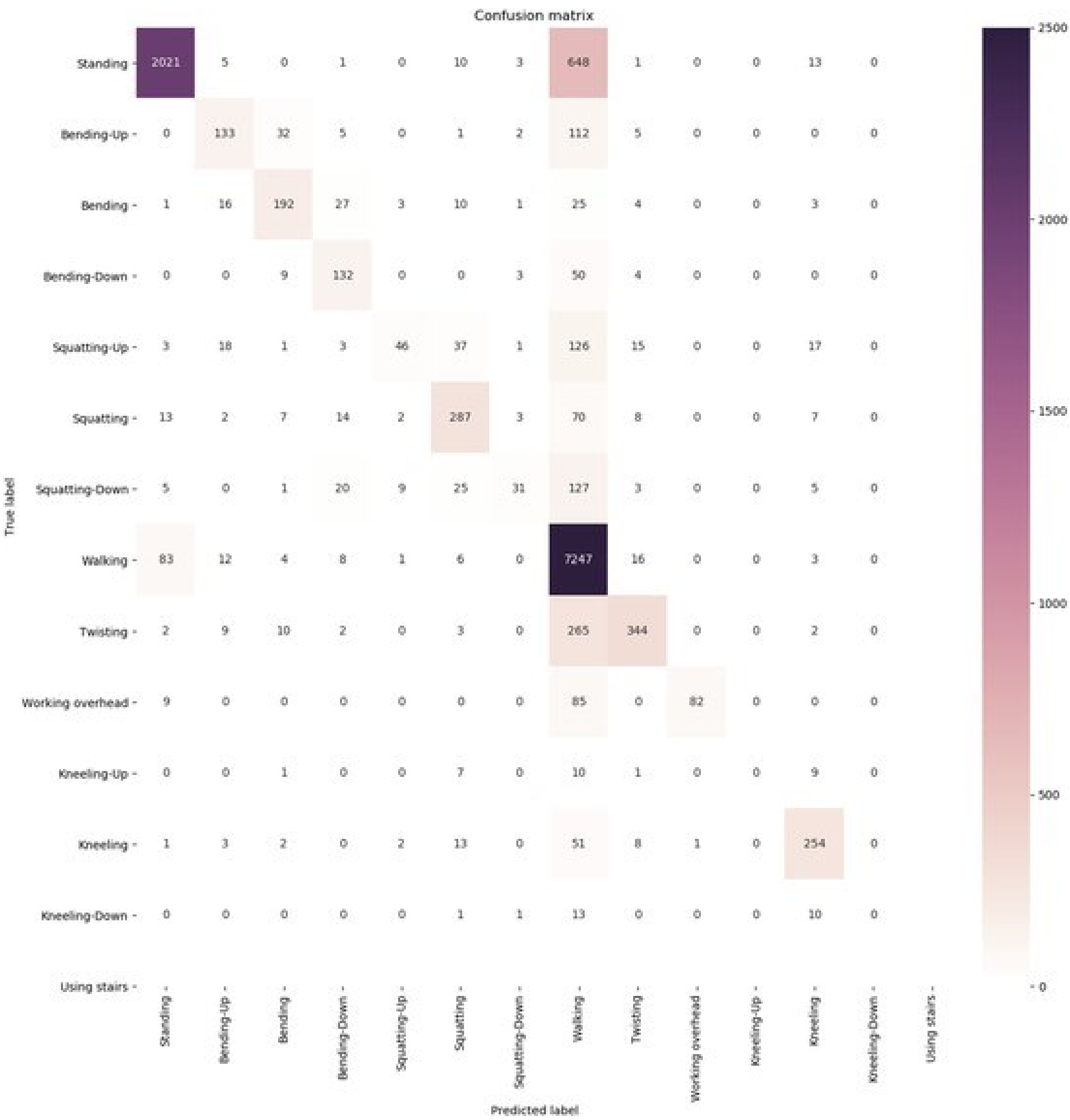

Fig. 9. Confusion matrix of the classification of the quasi-experimental test dataset.

Table 3
Hyper-parameters of the developed LSTM model.

Hyper-parameter	Value
The number of hidden units	64
L2 regularization factor	0.0008
Learning rate	0.0017
The number of epochs	400
Batch size	128
Dropout probability	0.2

fitting. Overfitting happens when the model is too closely fit the particular dataset and it shows low performance on the unseen data. To avoid the overfitting, l_2 norm regularization with the L2 regularization factor 0.0008 and a dropout technique were implemented. Dropout probability was set to 0.2 only in the training process. This means that the recurrent connections in LSTM cells were excluded with 20% of probability. The dropout technique was only deployed in the learning process on the train set. Adam optimizer [43] was used to minimize a loss function, i.e., the cross-entropy of the result after the softmax function was applied. As a result, the developed LSTM model showed the accuracy of 94.73% on the test set. Fig. 12 and Fig. 13 show the confusion matrixes of the result without normalization and with normalization, respectively. Kneeling-up and kneeling-down motion were omitted from the result because these motions were taken during a very short time that was less than 0.3 s. Since the label of the sequences was determined as the dominant label in the sequence, those two motions were omitted even if they were in the sequences.

To evaluate the developed LSTM model, precision-recall curves for each motion class were derived as shown in Fig. 14. Precision-recall (PR) curves are evaluation measures for classification that allows the visualization of the performance of the classifier at a range of thresholds [44]. PR curves are used to evaluate binary classification models trained using an imbalanced dataset where one class occupies a larger portion in the dataset than the other class. Since the dataset used in developing the LSTM model is an imbalanced dataset with multi classes, PR curves for each class are obtained. In the PR curves, the closer the curve is to a right upper corner, the better the classifier is. The PR curves for each class can be interpreted as a single value by calculating the area under the curve. The area ranges from 0 to 1, where 0 indicates the classifier completely failed to classify the data and 1 indi-

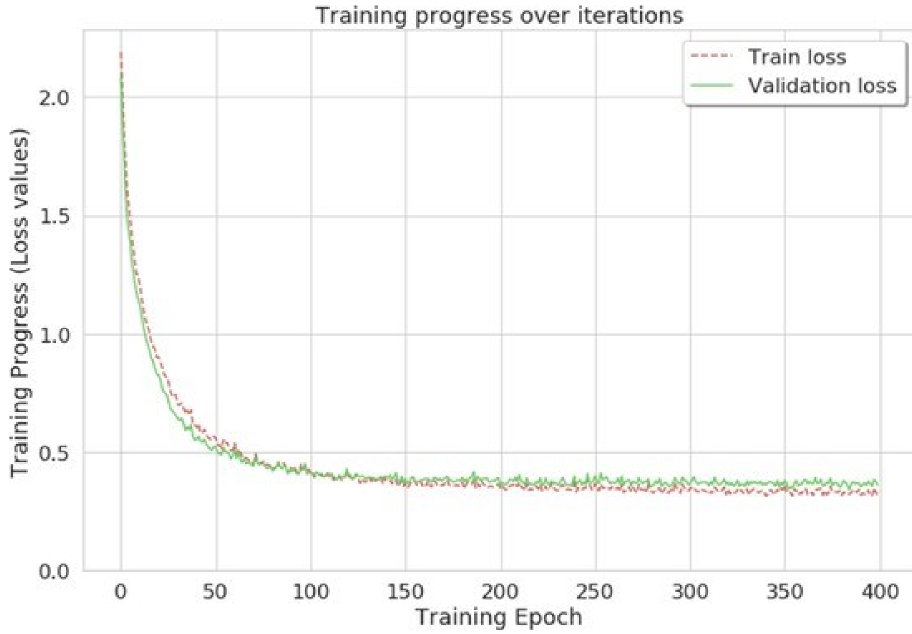

Fig. 10. Train and validation losses over iterations.

Fig. 11. Train and validation accuracy over iterations.

cates the classifier completely classified the data. The areas are shown in the legend of graphs in Fig. 14. The area of the averaged precision-recall for overall classes was 0.98. While the areas of the major motions including standing, bending, squatting, walking, twisting, working overhead, and kneeling were larger than 0.90, the areas of the transitioning motions including bending-up, bending-down, squatting-up, and squatting-down were smaller than those of the major motions.

5. Discussion

In the evaluation, various cases, where the different number of motion sensors located on different body parts were used to train machine learning classifiers, were investigated to identify the desired numbers and locations of the motion sensor for recognizing construction workers' motions. By comparing the accuracy of each case, the cases where two motion sensors located on the body parts in a certain distance showed the similar accuracy with the case where all motion sensors were utilized throughout the body. This indicates that the number of motion sensors required to recognize construction workers' motion can be reduced if their locations are selected based on the understanding of the effectiveness of the numbers and locations of the motion sensors. This finding was validated through the quasi-experimental test with multiple subjects. This is important because motion recognition systems can be efficiently designed with fewer sensors and still provide reliable performance in fields.

The target motion classes considered in the evaluation, quasi-experimental test, and development of LSTM model were 14 motions, which is more than one of the existing motion recognition models. In general, the more motions or activities that are considered, the

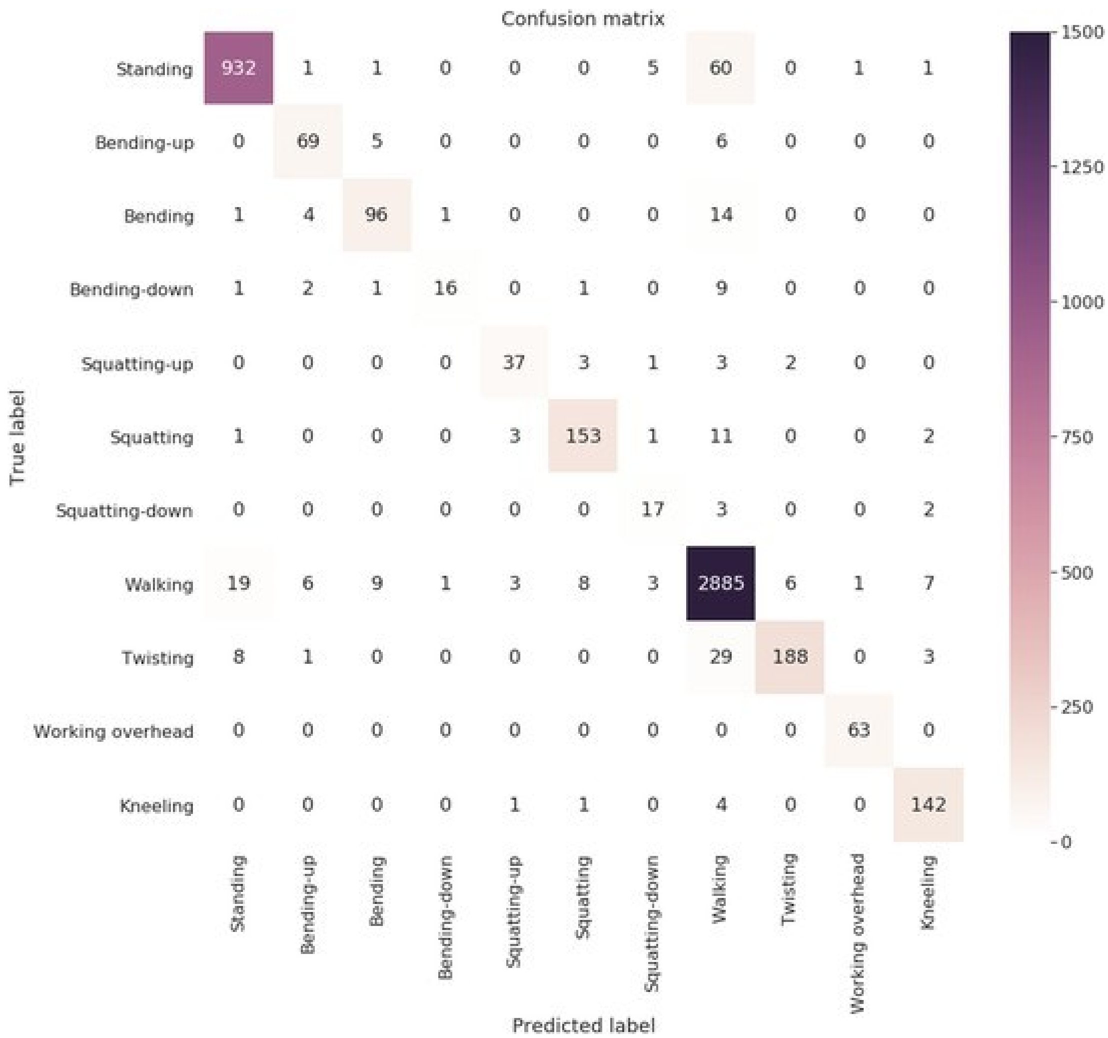

Fig. 12. Confusion matrix of the result of the LSTM model without normalization.

more difficult a classifier is trained or the larger the dataset is required. Hence, the current research efforts focus on recognizing particular motions of interest. In the proposed study, 13 motions were successfully recognized using the LSTM model with an accuracy of 94.73%. Compared to the best-performed accuracy of 82.39% among the five conventional machine learning algorithms, the developed LSTM network showed significantly higher accuracy. It was found that the developed model outperformed the existing motion and activity recognition methods in terms of accuracy, the number of classes, and the number of sensors as shown in Table 4. For example, compared to the convolutional LSTM [31], the developed LSTM model showed higher accuracy with fewer number of sensors and more number of target classes. This indicates that the developed model learned sequential patterns of the motions from the raw sensor data better than the one in which convolutional layers were adopted to the raw data.

The results of the quasi-experimental test show inherent motion distributions of the material handling tasks. For example, among 4853 sequential data points in the testing data, 2885 points were classified as 'Walking' motions. This indicates that the subjects spent more than 50% of their working time on carrying material or moving their position for the next material. Similarly, 932 points were classified as 'Standing' motions which mean the subjects were idling during about 20% of their working time. This distribution is naturally observed because carrying material is the most time-consuming activity in the material handling tasks while lifting or placing a material takes less time than carrying. Therefore, with the motion distributions of workers, their behavior can be individually monitored and analyzed.

The developed LSTM model recognized the major motions with very high accuracy. For example, the areas under PR curves of standing, walking, working overhead, and kneeling were over 0.99. This indicates that most instances in these motions were correctly classified. On the other hand, some of the motions such as bending-down and squatting-down motions showed smaller areas under PR curves than the other motions. It is because they are the transitioning motions, which are the motions between the major motions. The data corresponded to the transitioning motions contain less distinguishable patterns. However, the use of the transitioning motions has a strong advantage that allows the major motions to have more distinguishable patterns. Considering the applications of motion recognition models in construction projects, the major motions are the motions of interest for safety and productivity measurements. Therefore, the use of the transitioning motions can improve the practicality of the developed motion recognition model.

The successful implementation of the LSTM model was possible with the lower-performance computer because the size of the input data used in this study was much smaller than the one used in the existing efforts. Specifically, high-performance computers are required to process image data in the vision-based model because the model typically includes multiple convolutional layers and pooling layers to learn features from the images. However, in the case of the proposed

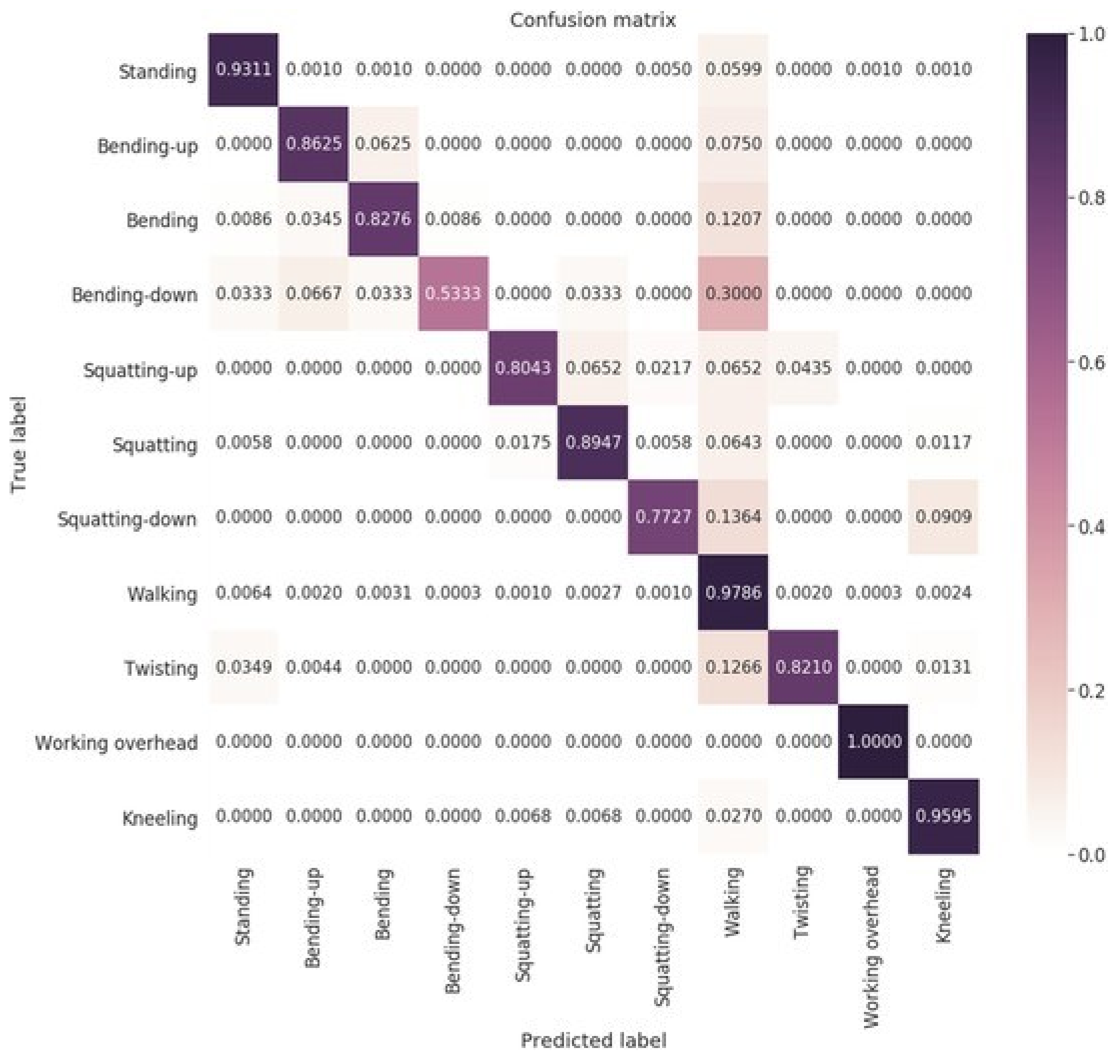

Fig. 13. Confusion matrix of the result of the LSTM model with normalization.

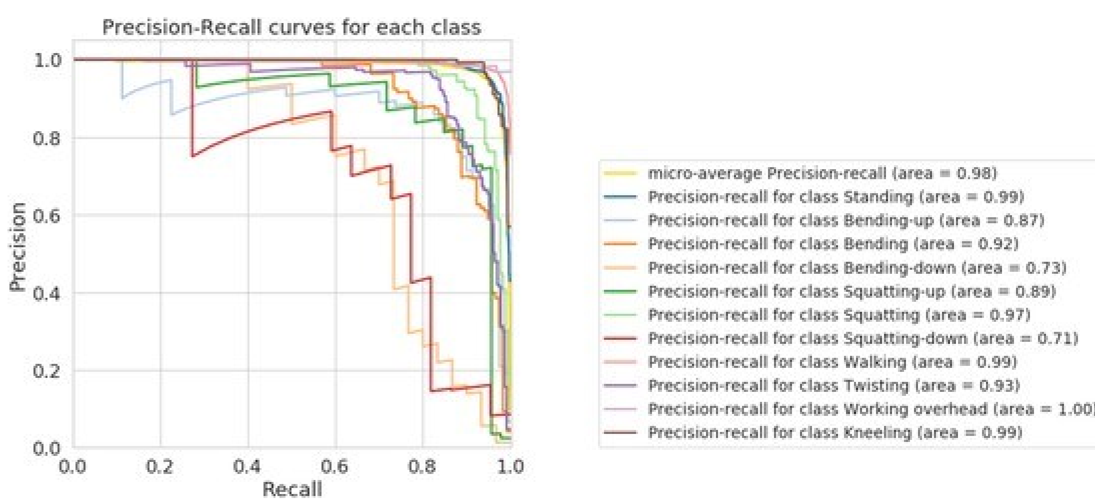

Fig. 14. Precision-recall curves for each motion class.

model, the dataset was a 16,178×18×40 matrix with float numbers. Thus, the model was able to be implemented without a high-performance computer.

The proposed study has several limitations. First, the datasets used in the evaluation, quasi-experimental test, and development of the LSTM model were not collected from real construction workers. The subjects involved in the study were students who are majoring in civil engineering. Hence, the dataset might not accurately represent construction workers' motions. Next, the datasets were imbalanced. Among 14 target motions, the walking motion occupied more than half of

Table 4
Comparison of the performance with existing methods.

Method	The number of sensors	Location(s) of the sensors	The number of classes	Accuracy
k-NN [21]	1	Arm	5	79.79%
Multi-class SVM [28]	1	Wrist	4	88.10%
Convolutional LSTM [31]	5	Head, arm, chest, thigh, and calf	8	85.20%
RF in this study	2	Hip and neck	13	82.39%
The developed LSTM in the this study	2	Hip and neck	11	94.73%

the entire dataset. This is because the subjects spent the most time on carrying the block. This imbalance might lead to a misestimation the performance of the model. Although the performance on each class was evaluated through PR curves, the overall performance of the model relied on the particular motions such as the walking and standing. Lastly, there is still a lack of the target motions in the dataset. While the datasets contain 13 or 14 motions that construction workers typically take during material handling tasks, construction tasks involve other motions that were not considered in this study. However, these limitations related to the dataset are expected to be overcome once more generalized and larger datasets are collected from actual construction workers.

6. Conclusion

This paper evaluated the effectiveness of the motion sensors' numbers and locations on construction workers' bodies and an LSTM model for recognizing their motions, which can be potentially used to measure workers' safety and productivity. In the evaluation, each of 32 different node combinations containing different numbers of motion sensors located on different body parts was used to train five machine learning classifiers. By comparing the performance of the classifiers, the desired numbers and locations of the motion sensors regarding the classification performance were identified. These findings were validated through the quasi-experimental test. To improve classification performance, the LSTM motion recognition model was developed. The same dataset used in the quasi-experimental test was used to develop the LSTM model. The model showed much higher classification performance on 13 motion classes than the conventional machine learning algorithms; the accuracy increased from 82.39% to 94.73%.

The main contribution of this study is twofold. First, this study provided an insight into the influences of the use of motion sensors in construction workers' motion recognition using a systematic approach. By generating multiple datasets containing motion sensor data collected from the sensors located on different body parts, the study investigated how the numbers and locations of motion sensors affect the performance of machine learning classifiers. Second, the LSTM network for recognizing construction workers' motion was presented. The LSTM network classifies 13 different construction workers' motions properly by learning sequential patterns in the motion sensor dataset. This model is expected to improve a construction worker monitoring mechanism for safety and productivity management by automatically identifying the workers' motions and working conditions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. (#1919068). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] CPWR – The Center for Construction Research and Training, The Construction Chart Book: The U.S. Construction Industry and Its Workers, 6th ed., 2018. https://www.cpwr.com/sites/default/files/publications/The_6th_Edition_Construction_eChart_Book.pdf(accessed July 17, 2019).
[2] The National Institute for Occupational Safety and Health (NIOSH) Ergonomics and Musculoskeletal Disorders, Workplace Safety & Health Topicshttps://www.cdc.gov/niosh/topics/ergonomics/default.html2019(accessed July 18,2019)
[3] McKinsey & Company, The Construction Productivity Imperative, McKinsey & Company, 2015. https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/the-construction-productivity-imperative(accessed January 22, 2018).
[4] P.R. Ghate A.B. More P.R. Minde Importance of measurement of labour productivity in constructionhttp://ijret.esatjournals.org201618 July 2019
[5] McKinsey Global Institute Reinventing construction: a route to higher productivityhttps://www.mckinsey.com/~/media/McKinsey/Industries/Capital Projects and Infrastructure/Our Insights/Reinventing construction through a productivity revolution/MGI-Reinventing-construction-A-route-to-higher-productivity-Full-report.ashx20171 February 2020
[6] L. Sveikauskas, S. Rowe, J. Mildenberger, J. Price, A. Young, Measuring productivity growth in construction, Monthly Labor Review (2018), doi:10.21916/mlr.2018.1.
[7] L. Pei, R. Guinness, J. Kaistinen, Cognitive phone for sensing human behavior, in: Z. Yan (Ed.), Encyclopedia of Mobile Phone Behavior, IGI Global, Hershey, PA, USA, 2015, pp. 1138–1150, doi:10.4018/978-1-4666-8239-9.ch093.
[8] S. Han, S.M. Asce, S. Lee, A.M. Asce, F. Peña-Mora, M. Asce, Vision-based detection of unsafe actions of a construction worker: case study of ladder climbing, J. Comput. Civ. Eng. 27 (2013) 635–644, doi:10.1061/(ASCE)CP.1943-5487.0000279.
[9] J. Yang, Z. Shi, Z. Wu, Vision-based action recognition of construction workers using dense trajectories, Adv. Eng. Inform. 30 (2016) 327–336, doi:10.1016/J.AEI.2016.04.009.
[10] L. Ding, W. Fang, H. Luo, P.E.D. Love, B. Zhong, X. Ouyang, A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory, Autom. Constr. 86 (2018) 118–124, doi:10.1016/J.AUTCON.2017.11.002.
[11] K. Kim, H. Kim, H. Kim, Image-based construction hazard avoidance system using augmented reality in wearable device, Autom. Constr. 83 (2017) 390–403, doi:10.1016/J.AUTCON.2017.06.014.
[12] H. Guo, Y. Yu, Q. Ding, M. Skitmore, Image-and-skeleton-based parameterized approach to real-time identification of construction workers′ unsafe behaviors, J. Constr. Eng. Manag. 144 (2018), doi:10.1061/(ASCE)CO.1943-7862.0001497 04018042.
[13] H. Zhang, X. Yan, H. Li, Ergonomic posture recognition using 3D view-invariant features from single ordinary camera, Autom. Constr. 94 (2018) 1–10, doi:10.1016/j.autcon.2018.05.033.
[14] J.A. Hess, L. Kincl, T. Amasay, P. Wolfe, Ergonomic evaluation of masons laying concrete masonry units and autoclaved aerated concrete, Appl. Ergon. 41 (2010) 477–483, doi:10.1016/J.APERGO.2009.10.003.
[15] H. Luo, C. Xiong, W. Fang, P.E.D. Love, B. Zhang, X. Ouyang, Convolutional neural networks: computer vision-based workforce activity assessment in construction, Autom. Constr. 94 (2018) 282–289, doi:10.1016/j.autcon.2018.06.007.
[16] V. Escorcia, M.A. Dávila, M. Golparvar-Fard, J.C. Niebles, Automated vision-based recognition of construction worker actions for building interior construction operations using RGBD cameras, Construction Research Congress 2012, American Society of Civil Engineers, Reston, VA, 2012, pp. 879–888, doi:10.1061/9780784412329.089.
[17] Y. Yu, H. Guo, Q. Ding, H. Li, M. Skitmore, An experimental study of real-time identification of construction workers′ unsafe behaviors, Autom. Constr. 82 (2017) 193–206, doi:10.1016/J.AUTCON.2017.05.002.
[18] S. Han, S. Lee, A vision-based motion capture and recognition framework forbehavior-based safety management, Autom. Constr. 35 (2013) 131–141, doi:10.1016/J.AUTCON.2013.05.001.
[19] A. Golabchi, S. Han, S. AbouRizk, A simulation and visualization-based framework of labor efficiency and safety analysis for prevention through design and planning, Autom. Constr. 96 (2018) 310–323, doi:10.1016/j.autcon.2018.10.001.
[20] J. Chen, J. Qiu, C. Ahn, Construction worker′s awkward posture recognition through supervised motion tensor decomposition, Autom. Constr. 77 (2017) 67–81, doi:10.1016/J.AUTCON.2017.01.020.
[21] R. Akhavian, A.H. Behzadan, Smartphone-based construction workers′ activity recognition and classification, Autom. Constr. 71 (2016) 198–209, doi:10.1016/J.AUTCON.2016.08.015.
[22] N.D. Nath, A.H. Behzadan, Construction productivity and ergonomic assessment using mobile sensors and machine learning, Computing in Civil Engineering, 2017, American Society of Civil Engineers, Reston, VA, 2017, pp. 434–441, doi:10.1061/9780784480847.054.

[23] R. Akhavian, A.H. Behzadan, Productivity analysis of construction worker activities using smartphone sensors, Proceedings of 16th International Conference on Computing in Civil and Building Engineering, 2016, pp. 1067–1074. https://pdfs.semanticscholar.org/9cc3/2a15b75765e977bc1b8ccc6d545e8c5168a2.pdf(accessed May 7, 2018).

[24] M.F. Antwi-Afari, H. Li, J. Seo, A.Y.L. Wong, Automated detection and classification of construction workers′ loss of balance events using wearable insole pressure sensors, Autom. Constr. 96 (2018) 189–199, doi:10.1016/j.autcon.2018.09.010.

[25] S. Xing, H. Tong, P. Ji, Activity recognition with smartphone sensors, Tsinghua Sci. Technol. 19 (2014) 235–249, doi:10.1109/TST.2014.6838194.

[26] J. Ryu, J. Seo, M. Liu, S. Lee, C.T. Haas, P.D. Student, Action recognition using a wristband-type activity tracker: case study of masonry work, Construction Research Congress, 2016, 2016, pp. 790–799. https://ascelibrary.org/doi/pdf/10.1061/9780784479827.080(accessed April 1, 2019).

[27] K. Yang, C.R. Ahn, Inferring workplace safety hazards from the spatial patterns of workers′ wearable data, Adv. Eng. Inform. 41 (2019), doi:10.1016/j.aei.2019.100924 100924.

[28] J. Ryu, J. Seo, H. Jebelli, S. Lee, Automated action recognition using an accelerometer-embedded wristband-type activity tracker, J. Constr. Eng. Manag. 145 (2019) 1–14, doi:10.1061/(ASCE)CO.1943-7862.0001579.

[29] K. Yang, C.R. Ahn, M.C. Vuran, S.S. Aria, Semi-supervised near-miss fall detection for ironworkers with a wearable inertial measurement unit, Autom. Constr. 68 (2016) 194–202, doi:10.1016/j.autcon.2016.04.007.

[30] A. Alwasel, A. Sabet, M. Nahangi, C.T. Haas, E. Abdel-Rahman, Identifying poses of safe and productive masons using machine learning, Autom. Constr. 84 (2017) 345–355, doi:10.1016/j.autcon.2017.09.022.

[31] J. Zhao, E. Obonyo, Convolutional long short-term memory model for recognizing postures from wearable sensor, CEUR Workshop Proceedings, 2019. http://ceur-ws.org/Vol-2394/paper36.pdf(accessed January 15, 2020).

[32] E. Valero, A. Sivanathan, F. Bosché, M. Abdel-Wahab, Analysis of construction trade worker body motions using a wearable and wireless motion sensor network, Autom. Constr. 83 (2017) 48–55, doi:10.1016/J.AUTCON.2017.08.001.

[33] K. Yang, C.R. Ahn, H. Kim, Validating ambulatory gait assessment technique for hazard sensing in construction environments, Autom. Constr. 98 (2019) 302–309, doi:10.1016/j.autcon.2018.09.017.

[34] H. Jebelli, C.R. Ahn, T.L. Stentz, Fall risk analysis of construction workers using inertial measurement units: validating the usefulness of the postural stability metrics in construction, Saf. Sci. 84 (2016) 161–170, doi:10.1016/J.SSCI.2015.12.012.

[35] H. Jebelli, S.M. Asce, Changbum, R. Ahn, A.M. Asce, T.L. Stentz, Comprehensive fall-risk assessment of construction workers using inertial measurement units: validation of the gait-stability metric to assess the fall risk of Iron workers, J. Comput. Civ. Eng. 30 (2015), doi:10.1061/(ASCE)CP.1943-5487.0000511.

[36] X. Yan, H. Li, A.R. Li, H. Zhang, Wearable IMU-based real-time motion warning system for construction workers′ musculoskeletal disorders prevention, Autom. Constr. 74 (2017) 2–11, doi:10.1016/j.autcon.2016.11.007.

[37] A. Singh, N. Thakur, A. Sharma, A review of supervised machine learning algorithms, proceedings of the 10th INDIACom, 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016, 2016, pp. 1310–1315. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7724478&isnumber=7724213.

[38] S. Dreiseitl, L. Ohno-Machado, Logistic regression and artificial neural network classification models: a methodology review, J. Biomed. Inform. 35 (2002) 352–359, doi:10.1016/S1532-0464(03)00034-0.

[39] S.-R. Ke, H. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, K.-H. Choi, A review on video-based human activity recognition, Computers 2 (2013) 88–131, doi:10.3390/computers2020088.

[40] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, Y. Amirat, Physical human activity recognition using wearable sensors, Sensors 15 (2015) 31314–31338, doi:10.3390/s151229858.

[41] F. Ordóñez, D. Roggen, Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition, Sensors 16 (2016) 115, doi:10.3390/s16010115.

[42] Y. Zhao, R. Yang, G. Chevalier, M. Gong, Deep residual Bidir-LSTM for human activity recognition using wearable sensors, Math. Probl. Eng. 2018 (2018) 13, doi:10.1155/2018/7316954.

[43] D.P. Kingma, J. Lei Ba, ADAM: a method for stochastic optimization, The 3rd International Conference for Learning Representations, San Diego, 2015. https://arxiv.org/pdf/1412.6980.pdf(accessed September 8, 2019).

[44] K. Boyd, K.H. Eng, C.D. Page, Area Under the Precision-recall Curve: Point Estimates and Confidence Intervals, Springer, Berlin, Heidelberg, 2013, pp. 451–466, doi:10.1007/978-3-642-40994-3_29.