Engineering Journal issn: 0125-8281

Article

Development of Automated System for Classifying Productivity Behavior of Construction Workers Using Deep Learning

Ryuji Kasai[1,a], **Takashi Goso**[2,b,*], **Mizuki Akiyama**[1,c], **and Hiroto Kato**[1,d]

1 Tokyo City University, Tokyo, Japan
2 Faculty of Architecture and Urban Design, Tokyo City University, Tokyo, Japan
E-mail: [a]g2281616@tcu.ac.jp, [b,*]tgoso@tcu.ac.jp (Corresponding author), [c]g2381601@tcu.ac.jp, [d]g2032024@tcu.ac.jp

Abstract. In Japan, the integration and comprehensive understanding of data related to the working environment and productivity at construction sites remain underdeveloped. This study introduces a system that utilizes the human activity recognition method, employing accelerometers combined with deep learning techniques, to capture a detailed overview of activities performed by construction site workers. We developed a new approach for transforming accelerometer data collected from devices attached to workers' helmets into a format suitable for image-based analysis. This data was then processed using a convolutional neural network to create a deep-learning model capable of distinguishing between different types of worker movements. The model demonstrated high accuracy, with correct classification rates of 80.0% for walking and 92.1% for forward-leaning postures—activities commonly observed at construction sites. Additionally, we established an ensemble system to enhance the final classification of productive motions. This innovative system holds the promise of enabling future quantification of on-site productivity through daily indices that reflect workers' engagement levels.

Keywords: Worker productivity data, work sampling, deep learning, acceleration, convolutional neural network.

ENGINEERING JOURNAL Volume 28 Issue 10
Received 7 March 2024
Accepted 17 October 2024
Published 31 October 2024
Online at https://engj.org/
DOI:10.4186/ej.2024.28.10.59

DOI:10.4186/ej.2024.28.10.59

1. Introduction

This study aims to develop an innovative system designed to enhance productivity and workflow optimization on construction sites. This system, utilizing advancements in sensing technology and deep learning, aims to merge traditional understandings of manual productivity with the human activity recognition (HAR) method. It seeks to provide a comprehensive assessment of construction workers' activities, enabling the quantitative evaluation of worker efficiency upon the introduction of new tasks. Furthermore, this system aims to facilitate process and workflow reviews and support the implementation of new work types.

Central to this initiative is the creation of a tool that not only quantitatively demonstrates the system's efficacy in introducing new work processes but also integrates seamlessly with sensors mounted on workers' helmets. These sensors, strategically positioned around the head, are pivotal in evaluating the system's potential as a foundational component of an occupational safety management system. This system would be capable of aggregating critical information to prevent workplace accidents and ensure rapid response to any incidents.

For the system to effectively serve as a productivity enhancement tool, it must autonomously record data with consistent accuracy and continuity without necessitating human intervention. The data collected include the duration of each task performed by on-site workers, locational dynamics within the workplace, and acceleration metrics determined through an inertial measurement unit (IMU) affixed to the worker's helmet. This research aims to develop a system capable of discerning actionable insights via deep learning from such data. These insights include work environment optimization, work intensity adjustments for each worker, accident prevention strategies (notably those involving construction machinery and falls), and the identification of productivity impediments.

2. Targeted Construction Types and Occupations

This study specifically examines outdoor pavement construction, characterized by sequential operations and the directional movement of workers over time. This study specifically examines outdoor pavement construction, characterized by sequential operations and the directional movement of workers over time. The reason for targeting this type of work is mainly because it is an outdoor work where location information can be obtained at a certain level. Another reason is that the work area is continuous and the workers do not move in a multi-level intersection, making observation easy. Furthermore, since these projects are all paving work and the tasks are routine, primarily using construction equipment, it was considered easy to classify the types of work. The segmentation of construction tasks and the variability in workers' presence—some being on-site for only a single day—pose significant challenges for data collection. Additionally, the use of global navigation satellite system (GNSS) technology is impractical for tasks such as tunnel lining, complicating the acquisition of location data in environments with restricted areas or where manual work sampling is impeded.

Outdoor pavement construction is distinguished by its repetitive sequence of activities, including the spreading of pavement by base pavers followed by compaction using vibratory and tire rollers. The relatively slow progression of these tasks and the absence of overhead obstacles uniquely position outdoor construction sites for the study. Such conditions are anticipated to result in infrequent detections of irregular acceleration, simplifying the analysis of movement across different job categories. Workers in pavement construction undertake various roles, from operating heavy machinery and compacting materials with rakes to manual tasks involving shovels. The diversity of these roles further highlights the complexity of accurately classifying and analyzing work patterns.

The methodology encompasses the entire pavement construction process, from the movement of roadbed and base pavers in a single direction—with materials fed from the direction of travel—to the formation of the road behind the equipment. This process involves multiple job functions beyond equipment operation, including assistant operators managing ancillary machinery and workers adjusting the spread of materials. Data sampling efforts were extended across these varied occupations to ensure a comprehensive understanding of productivity and workflow within outdoor pavement construction.

3. Previous Research

In Japan, surveys related to construction and labor costs are conducted by organizations such as the Construction Research Institute, which operates under contracts from various client entities, including ministries, local governments, independent administrative bodies, and corporations. These surveys document numerous

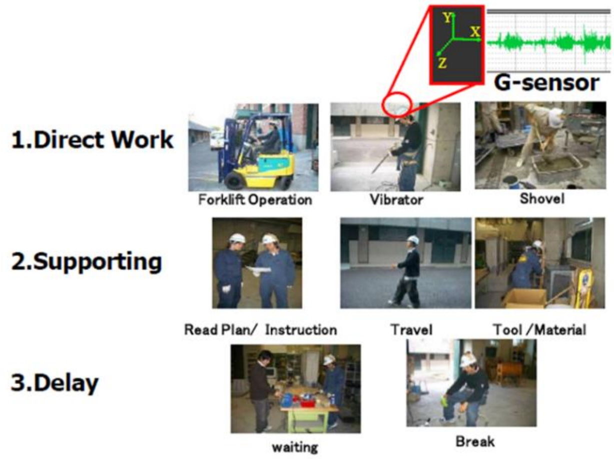

Fig. 1. Image of classification of worker's behaviors at construction sites.

details, such as quantities of construction materials, workforce numbers, types of jobs, hours worked, and the usage of machinery and equipment based on standardized forms. The primary goal is to accurately capture the labor conditions and physical productivity (work rates) at different construction sites.

In the United States, efforts to gather data on construction worker productivity through work-sampling surveys have been underway since the 1980s. Surveyors visit sites to observe and categorize the activities of workers into pre-established classifications, enabling an analysis of site-specific characteristics and their evolution over time. This approach, already prevalent in Japan's manufacturing sector, was further refined for construction settings by Kusayanagi [1], who developed categories for work operations as illustrated in Table 1.

Direct production activities refer to tasks directly related to productivity. Auxiliary support activities comprise tasks that, while not directly contributing to productivity, facilitate or support productive actions. Work delay activities are those that hinder or delay the progress of both direct production and auxiliary support activities. This categorization is visualized in Fig. 1. However, the practicality of implementing this method, which demands extensive visual monitoring of work status, as a routine management tool in construction sites is questionable. This is due to its significant demands on time, specialized knowledge, staffing costs, and adaptability to the dynamic nature of construction sites, where work tasks and personnel often change daily. These challenges likely explain the method's limited adoption in Japan's construction sector.

Table 1. Classification of productivity behaviors and their components at construction sites.

Productivity Behavior Category	Work Component Category
Direct work	0. Direct work
Support	1. Verification of drawings and instructions
	2. Worker movement
	3. Transport of materials and equipment
	4. Preparation of tools and materials
Delay	5. Work commencement delays
	6. Waiting/standby
	7. Delays due to personal reasons
	8. Breaks

3.1. Methods for Grasping and Monitoring Activities in Construction Industry

Previous efforts to develop motion detection and monitoring systems in the construction sector have varied significantly in terms of methods and objectives as described below. Three primary monitoring approaches have been identified: computer vision-based, audio-based, and kinematic-based methods. Motion recognition predominantly relies on wearable sensors for sampling and utilizes positional data to track movement patterns. These monitoring systems are generally designed to enhance occupational safety and enable uncrewed operations across construction sites.

In domestic contexts, one notable application of wearable sensors involves the precise detection of specific work activities, such as rebar tying, by analyzing motion data from an accelerometer attached to the worker [2]. This method demonstrates promising potential for high-accuracy activity recognition in focused tasks.

Video analysis techniques have also been employed, utilizing footage from fixed-point cameras to identify materials and equipment on-site. By integrating this visual data with the weight and location information of materials, researchers can estimate the configuration of workpieces and assess physical labor productivity [3].

Furthermore, there are systems designed to mitigate industrial accidents by alerting users to the proximity between workers and cranes. These systems integrate Global Positioning System (GPS) data with three-dimensional computer-aided design models to monitor potential safety hazards [4]. Despite various monitoring techniques used within the i-Construction movement, comprehensive site monitoring for productivity enhancement remains relatively underexplored, indicating a need for further technological development in this area.

3.2. Background of Productivity Behavior Classification System Development

3.2.1. ZigBee-Based System by Goso, Ochi, and Kusayanagi

Goso et al. [5-8] developed a specialized device utilizing a three-axis accelerometer and ZigBee technology to prototype a standard for distinguishing productivity categories. Their methodology involved analyzing vibration patterns from accelerometer data attached to workers to categorize productivity levels. Challenges remaining in their research include refining the criteria for productivity classification programming and efforts to minimize both the size and the cost of the hardware.

3.2.2. Productivity Behavior Classification System via Pre-Classification and Machine Learning by Osawa and Goso

Osawa et al. [9-11] attached commercially available three-axis accelerometers, GNSS sensors, and thermo-hygrometers to workers, integrating acceleration and positional data. Through the application of support vector machine (SVM), fast Fourier transform (FFT) analysis, and other techniques, they developed a methodology that integrates binary comparisons of specific criteria, such as variability and periodicity, to classify productivity-related movements. This approach successfully demonstrated that productivity movements could be categorized with a reasonable level of accuracy by utilizing acceleration data and positional information. However, the system's reliance on manually set threshold values for classification and the inclusion of subjective judgment in these thresholds present significant limitations. Specifically, it is more challenging the application of this system to different work types or to the same work types under varying site conditions. Additionally, the approach faces challenges in maintaining time-series data integrity, as statistical processing of acceleration data over certain periods—such as calculating mean and variance—for labeling purposes does not preserve temporal information.

3.3. Research on HAR in Other Fields[12-49]

HAR represents a collection of technologies designed for the automatic detection and classification of human activities. Its applications span healthcare, sports, security, and ubiquitous computing, reflecting its broad utility and relevance.

HAR methodologies primarily incorporate machine learning and pattern recognition techniques. These approaches involve extracting features from sensor-derived data, which are then analyzed by a learning algorithm to differentiate between various activities. The effectiveness of HAR systems depends significantly on the volume of data available and the precision of feature selection, highlighting the critical role of comprehensive datasets and accurate feature identification. Recent studies frequently utilize a single IMU or a combined device setup.

Since 2010, the advent of deep learning methodologies, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has advanced HAR research. These techniques facilitate the extraction of complex activity patterns and enhance the system's resilience against data variability, thereby enabling real-time activity recognition and significantly improving both usability and safety.

In the domain of sports, for instance, integrating tracking sensors attached to soccer players' footwear (in the English Premier League) with ball trajectory data and stationary camera footage has led to the development of models capable of analyzing player activities with remarkable precision. Such models can accurately assess the dynamics of team sports, including pass counts, shot attempts, movement intensity, and player positioning relative to the ball [16].

Another innovative application involves using accelerometer-equipped shoes to automatically identify and classify breakdance movements into 17 distinct categories. This technology not only evaluates the originality of dance sequences but also provides invaluable feedback to performers [17].

In Bangladesh, a case study of HAR research within the nursing care sector primarily targeted motion recognition for rehabilitation, elderly care, and the prevention of in-facility accidents. Utilizing long short-term memory (LSTM)networks combined with CNNs, researchers achieved high-accuracy motion detection by correlating IMU data from smartphones with mHealth data, highlighting the potential of HAR in enhancing patient care and safety [18].

Security applications of HAR focus on the identification of suspicious behaviors, indicating the technology's potential to contribute to safer environments. The ongoing development of sophisticated HAR techniques promises to extend its applicability to increasingly diverse and complex scenarios.

3.4. Research on Time-Series Data Classification Using Deep Learning

The study by Osawa et al. encountered limitations in retaining time-series information from acceleration data. To address this, we explored deep learning methodologies capable of classifying data while preserving time-series information, including CNNs, LSTM networks, transformers, and gradient-boosting decision trees (GBDTs). CNNs and LSTMs, in particular, have gained prominence in accelerometer-based HAR due to their proficiency in processing multidimensional information, which includes handling three-axis acceleration and temporal data, as demonstrated in the Bangladesh study mentioned earlier. The efficiency in tensor computation these algorithms offer made them the chosen methods for this investigation.

3.4.1. Overview of CNN

CNNs represent a class of deep learning models recognized for their exceptional performance in image recognition and pattern detection tasks. These models have driven significant advancements in image processing through a foundational structure comprising convolutional layers, pooling layers, and fully connected layers. The convolutional layer extracts image features, the pooling layer improves computational efficiency by reducing the feature size, and the fully connected layer performs class classification. The convolutional layer performs convolution operations, where a kernel or filter matrix is applied to the input data to produce feature maps. This operation is mathematically represented as follows:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n) \cdot K(i-m, j-n), \quad (1)$$

where $S(i,j)$ denotes an element in the output feature map, I is the input data, and K represents the kernel matrix. This process applies the convolution operation across the input data using sliding filters to generate comprehensive feature maps.

The pooling layer follows, aimed at reducing the dimensionality of the convolutional layer's feature maps to enhance computational efficiency. Max pooling, a prevalent technique, constructs new feature maps by

DOI:10.4186/ej.2024.28.10.59

selecting the maximum value within specific regions, described as follows:

$$P(i,j) = \max_{m,n} S(i \times s + m, j \times s + n), \quad (2)$$

where $P(i,j)$ is the pooling layer's output element, S is the convolution layer's output, and s is the stride. This operation reduces feature map sizes while preserving essential spatial information.

Subsequent to the convolution and pooling layers is the fully connected layer, where every neuron from the previous layer is linked to every neuron in the next layer, resulting in the output layer. Typically, a SoftMax function computes the probability distribution over various classes, facilitating the network's ability to predict the likelihood of the input data belonging to each class.

$$y_k = \frac{e^{z_k}}{\sum_j e^{z_j}}, \quad (3)$$

where y_k represents the probability associated with class k, and z_k denotes the score for class k given the input vector.

3.4.2. Overview of LSTM

LSTM networks, a specialized form of RNNs, are designed to address and overcome the limitations of traditional RNNs in capturing long-term dependencies within time-series data. A distinctive feature of LSTM networks is their ability to mitigate the vanishing gradient problem, enabling the effective learning and retention of information over extended sequences. The core architecture of an LSTM includes a cell state and a series of gates that regulate the flow of information.

The input gate, which introduces data into the cell state, operates according to the following equations:

$$i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}), \quad (4)$$
$$\tilde{C}_t = \tanh(W_{ic}x_t + b_{ic} + W_{hc}h_{t-1} + b_{hc}), \quad (5)$$

where i_t is the input gate's output, x_t is the current input, h_{t-1} represents the hidden state from the previous timestep, W denotes the weight matrices, b is the bias, σ is the sigmoid activation function, and $tanh$ indicates the hyperbolic tangent function.

The forget gate, which determines the extent to which previous information is retained or discarded, is defined by

$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}). \quad (6)$$

From this information, the cell state is updated as follows:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t. \quad (7)$$

The final output is then

$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}), \quad (8)$$
$$h_t = o_t \cdot \tanh(C_t), \quad (9)$$

where o_t is the output gate's activation, and h_t is the resultant hidden state. This output is modulated by the output gate and then passed through a tanh function to produce the final hidden state. Through these mechanisms, LSTMs efficiently capture and model long-term dependencies in time-series data, significantly addressing the challenges posed by the gradient vanishing problem. In this study, we developed a system capable of analyzing productive motion through acceleration data while preserving time-series information. This was achieved by employing both CNN and a hybrid CNN+LSTM approach, wherein acceleration data across three axes was processed using the methodologies described below.

4. Construction Site-Specific Work Data Acquisition and Training Data Creation

The methodology for data collection and training is detailed in this section.

4.1. On-Site Data Acquisition Environment

Data collection was conducted at an active pavement construction site to capture acceleration data associated with paving activities. The setup for sensor installation and data sampling is depicted in Figs. 2–4. The procedural steps undertaken are outlined as follows:

- Sensor and smartphone setup: Accelerometers and smartphones were powered on and synchronized with the current time.
- Bluetooth connection: Devices were paired via Bluetooth to ensure seamless data transmission.
- Acceleration data verification: The application linked to the accelerometer was checked to confirm the capture of acceleration data.
- GPS functionality check: The smartphone's GPS function was activated and checked for data acquisition and storage capabilities.

Fig. 2. Installation of sensors on helmets and connection status via Bluetooth.

DOI:10.4186/ej.2024.28.10.59

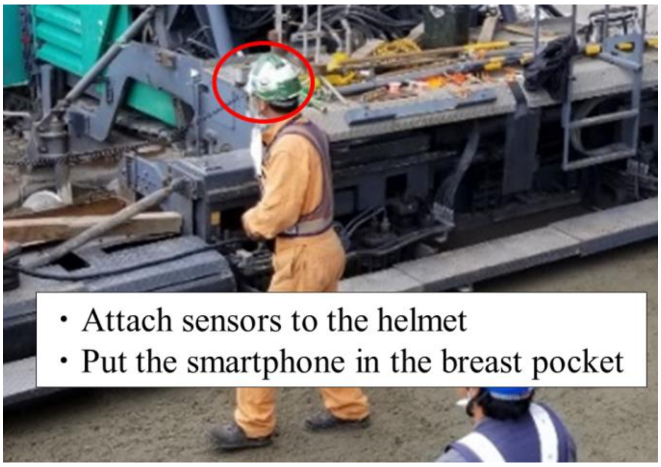

Fig. 3. Worker wearing sensors.

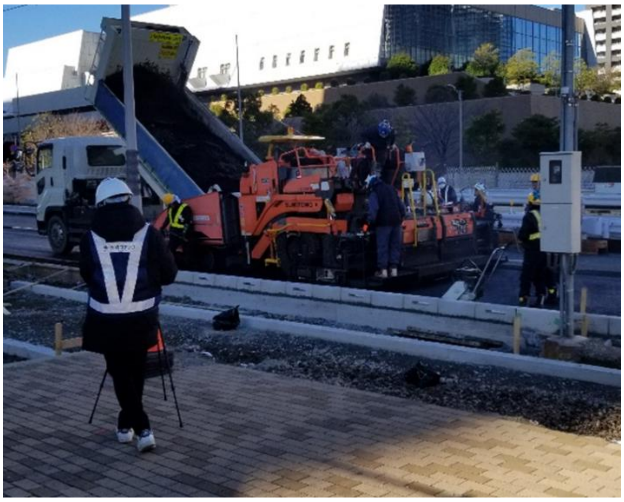

Fig. 4. Research team filming the construction work.

- Data logging application: The accelerometer and GPS logging application were initiated to begin data collection.
- Worker engagement: Workers were informed about the data collection initiative, highlighting that the objective was purely for research and not for monitoring work performance.
- Sensor attachment to helmets: Workers' helmets, five in total, were equipped with accelerometers using duct tape for secure attachment, as illustrated in Fig. 2. Different colors of duct tape facilitated worker identification.
- Smartphone carriage: Workers were requested to carry a smartphone, as shown in Fig. 3, to complement the acceleration data with location information.
- Video documentation: Concurrent with the commencement of work activities, researchers recorded the tasks performed by assigned workers (Fig. 4). This footage was intended for subsequent manual annotation of the acceleration data to serve as training data.

We developed these procedures after several improvements based on our experience with failures such as the inability to acquire sensor data or disconnection between the sensor and smartphone, which occurred through multiple rounds of data sampling. Although empirical, we have summarized the sampling method in our laboratory and prepared a manual. A significant focus was on capturing data during the combined processes of roadbed and base layer construction. Efforts were made to consistently film from perspectives that maintained the relative positions of heavy equipment operators and other workers, facilitating accurate work data acquisition. This approach is based on the understanding that basic movements such as standing, sitting, walking, and running—common categories in HAR—are partially discernible through these methods. Furthermore, location data is reliably obtained via satellite for work conducted outside of tunnels or mountainous regions.

The WT901BLECL (Shenzhen Wit Intelligent Co., Ltd.) served as the IMU with a three-axis accelerometer paired with the moto e7 (Motorola Mobility LLC.) for data logging. While not a focus of this study, temperature and humidity were also recorded using the IBS-TH2 PLUS (Shenzhen Inkbird Technology Co., Ltd.), considering potential applications in occupational safety management systems. Attaching the device to the helmet added approximately 82 g to its weight, a negligible increase when compared to the 100–120 g addition from standard field equipment such as face shields or headlights. This minimal weight adjustment, alongside strategic placement to avoid lateral shifts, ensured that the helmets remained comfortable and unobtrusive.

The IMU's sampling rate was set at 50 Hz, enabling the capture of 50 data points per second across the X, Y, and Z axes. This setup facilitated the generation of accurate labels for five basic actions, further detailed in the following sections.

4.2. Construction Site-Specific Behavior and Productivity Behavior Classification

Prior foundational studies in HAR have predominantly focused on categorizing simple, repetitive motions such as standing, sitting, walking, and running. However, the pavement construction activities addressed in this research encompass a broader spectrum of complex tasks, each integrating these basic movements within diverse operational contexts on construction sites. Applying traditional HAR methods to capture the characteristics of these tasks proves inadequate. For instance, raking activities might superficially resemble walking or be a subset thereof when analyzed through conventional accelerometer-based HAR techniques. Yet, the variability in an individual's physique, raking habits, and the nonuniform application of the tool itself mean that identifying raking as merely walking falls short of accurately classifying the task. Moreover, relying solely on simple actions such as walking makes it difficult to clearly distinguish between productivity-enhancing movements and those without direct contributions to work output. This includes differentiating between tasks such as ground clearing (walking with work), soil leveling via heavy machinery (work without walking), and nonproductive movements such as moving from the site to the office (walking without work).

These considerations suggest that the action classifications used in traditional HARs are insufficient for directly correlating with the three categories of productivity actions and the nine work component categories outlined in Table 1, specific to construction site operations.

4.3. Issues Related to Data Preparation and Definition of Behavior in this Study

4.3.1. Osawa's Classification

In Osawa's prior research, the methodology was devised to accurately classify movements into distinct categories of productive behavior, in alignment with Kusayanagi's framework for productivity-based movement classification. This approach integrated multiple mathematical techniques to correlate specific movements with the intended productivity categories. The attributes considered for classification included the presence/absence of variability, presence/absence of stationarity, average acceleration value, acceleration standard deviation, posture angle (pitch angle), and acceleration FFT analysis. To achieve high classification accuracy, threshold values had to be set for each dataset. However, this method presented challenges in maintaining consistent classification criteria across different types of work, even when the workers' actions were nearly identical. The necessity for subjective determination of these criteria by analysts introduced a layer of complexity, compromising the ability to ensure stable and universally applicable accuracy standards.

4.3.2. Combining Five Simple Actions

In this study, addressing the challenges inherent in the productivity motion classification systems developed by Goso et al. and Osawa et al., our objective was to preserve time-series information while enhancing the reliability of the classification process. We aimed to minimize subjective judgments by automating the classification process as much as feasible. To reduce the impact of subjective bias, we used objective data derived from analyzing motion characteristics common across construction site tasks, utilizing video footage collected during the creation of the training dataset. This footage, encompassing over 60 h of labor, including roadbed and base layer work, was instrumental in identifying movement patterns applicable across various work types while aligning with productivity movement classifications.

Initially, we simplified movement analysis to two primary questions: "Is work being performed?" and "Is walking occurring?" We defined "work" as movements of the upper limbs and torso that are deliberate and exceed a predefined threshold, excluding involuntary physiological actions such as yawning or sneezing. Additionally, activities characterized by stationary checking behaviors without significant upper limb or torso movements were classified as "no work." Most "with work" movements were identified by variations in upper limb and torso acceleration. However, verification actions, similar to those observed in the worker depicted in Fig. 3, were classified as "no work" due to their nature of momentarily halting work that involves upper limb or torso movements.

We applied these work and walking criteria in a preliminary experiment to assess the feasibility of using a CNN for classification. As detailed in Section 5, this approach enabled us to distinguish between the presence and absence of walking and work with an approximate accuracy of 80%.

In addition to identifying work and walking movements, classifying nonproduction movements in construction work based on their presence alone proves challenging. To address this, we established five simple

Table. 2 Five simple actions and their identification criteria.

Simple-Action Item	Criteria for Classification
Walking or not walking (walk)	Presence of walking within a specified time range; no distinction between walking for moving, carrying, and working.
With or without forward leaning (bent)	Identification of forward-leaning (with the neck pointing downward) based on a tilt of more than 10°.
Periodic and regular movements (repetitive motion)	Actions with a regular period, such as using shovels, rakes, or rammers, expected to show constant acceleration.
No periodicity and distinctive behavior (characteristic motion))	Direct production activities and others not expected to be periodic, indicating intermittent but identifiable work-related movements.
Considerable movement, although not contributing to classification (ineffective motion)	Movements unrelated to productivity classification, such as physiological actions (e.g., yawning, sneezing) that could introduce noise into the classification.

actions that could influence classification under the assumption that even complex tasks contain characteristic elements. These actions are detailed in Table 2, including walking, bending, repetitive motion, characteristic motion without periodicity, and ineffective motion (which does not aid in classification).

The criteria for walking involved assessing the time span and the context of walking, distinguishing between movement for support activities (e.g., moving, carrying) and direct production activities. The bent category was determined by the presence of a forward-leaning posture, identified by a neck or upper body tilt of approximately 10°. However, refining these criteria could enhance accuracy, a topic revisited in Section 6.2 for future research.

In addressing the category of repetitive motion, we classified actions characterized by consistent, short cycles, such as the use of shovels, rakes, and rammers commonly employed in pavement construction. This category was broadened to include a wide range of construction activities, including concrete pouring and rebar installation. Additionally, it covers movements anticipated to induce minor accelerations due to the operation of construction machinery and equipment, such as engaging with and maneuvering heavy machinery or vibrating concrete.

For the characteristic motion without periodicity, we identified movements that either lacked a regular time interval, indicative of repetitive motions or involved significant upper body acceleration variations, particularly in continuous tasks. This category was primarily aimed at identifying direct production activities, distinguishing them through notable acceleration shifts, which are uncommon in motions related to ancillary support or work delays, as observed in on-site video analysis.

The ineffective motion category was designed to isolate significant movements derived from physiological phenomena occurring over extended periods on construction sites, which do not contribute to the classification of productive activities. Previous research often misclassified any substantial acceleration changes as directly productive motions. However, transient acceleration spikes due to sneezing, coughing, yawning during breaks, or stretching to relieve tension were observed. While infrequent, such movements could skew classification outcomes. This category serves to identify and classify these as noncontributory noise.

Given the demonstrated efficacy of CNNs and LSTM networks in achieving high accuracy in binary and focused motion classification, our approach utilizes these models for binary classification based on the five simple actions, utilizing the same acceleration data. This method aims to train the models to recognize five distinct features of motion within the same timeframe.

4.4. Differences from Existing HAR Training Data and Handling Issues

In existing HAR methodologies, the generation of training data typically involves the repetition of identical motions for a specified duration, monitored through velocity meters affixed to various body locations to capture precise motion data. However, such sampling techniques prove infeasible in construction site settings, where constant motion is rare and the work environment evolves as projects progress. Consequently, this study required manual, second-by-second sampling based on visual observations and recorded data, necessitating the labeling of movements. A significant challenge encountered was the difficulty in uniformly collecting data over extended sampling periods, leading to datasets with considerable variance. This variance limited the verification of some movements due to insufficient training data. Nonetheless, a system framework capable of aggregating this data was developed, anticipating its future utility as the volume of samples expands. Foreseeing the substantial data variance, the dropout method was applied to mitigate overfitting, and data from multiple workers were randomized and trained while preserving temporal sequences to effectively minimize bias.

4.5. Creation of Training Data

To minimize personal bias, we established specific criteria for generating the training data derived from the guidelines in Table 2 and actual video footage. This process was validated with the assistance of students from our laboratory. The training data were segmented into intervals as brief as 1 s, with each segment containing five accurate labels, indicating the importance of synchronizing these labels with the corresponding acceleration data. Consequently, the timing between the smartphone used for logging and the data captured for label generation was regularly verified during on-site sampling to ensure consistent alignment across all datasets. Furthermore, given the temporal nature of pavement construction activities, we utilized multiple cameras to capture images from various angles, periodically repositioning these devices to mitigate any coverage gaps.

In the process of generating accurate labels, video footage was reviewed by several individuals. Labels were assigned to each of the five categories at 1 s intervals, taking into consideration the influence of preceding and subsequent movements. Given that the system was developed using Python, the data was formatted as .csv files for ease of manipulation. An excerpt of the training data is illustrated in Fig. 5.

5. Development of Productivity Operation Systems

5.1. Issues with Existing System

The existing classification system, which utilizes mathematical methods, achieved a classification accuracy of nearly 70%. However, this system required manual tuning for each type of work, including the adjustment of threshold values and specific calculations for each work type. While binary classification via SVM could handle

DOI:10.4186/ej.2024.28.10.59

1	動画10035	time	category	activity	walk	rep-motion	char-motion	ineff-motion	bent
620	14分39秒	9:31:43	0	0	0	1	0	0	1
621	14分40秒	9:31:44	0	0	0	1	0	0	0
622	14分41秒	9:31:45	0	0	0	1	0	0	0
623	14分42秒	9:31:46	2	6	0	0	0	0	0
624	14分43秒	9:31:47	2	6	0	0	0	0	0
625	14分44秒	9:31:48	2	6	0	0	0	0	0
626	14分45秒	9:31:49	2	6	0	0	0	0	0
627	14分46秒	9:31:50	2	6	0	0	0	0	0
628	14分47秒	9:31:51	2	6	0	0	0	0	0
629	14分48秒	9:31:52	2	6	0	0	0	0	0
630	14分49秒	9:31:53	2	6	0	0	0	0	0
631	14分50秒	9:31:54	2	6	0	0	0	0	0
632	14分51秒	9:31:55	2	6	0	0	0	0	0
633	14分52秒	9:31:56	2	6	0	0	0	0	0
634	14分53秒	9:31:57	2	6	0	0	0	0	0
635	14分54秒	9:31:58	2	6	0	0	0	0	0
636	14分55秒	9:31:59	2	6	0	0	0	0	0
637	14分56秒	9:32:00	2	6	0	0	0	0	0
638	14分57秒	9:32:01	2	6	0	0	0	0	0
639	14分58秒	9:32:02	2	6	0	0	0	0	1
640	14分59秒	9:32:03	0	0	0	1	0	0	1
641	15分00秒	9:32:04	0	0	1	1	0	0	1
642	15分01秒	9:32:05	0	0	1	1	0	0	1
643	15分02秒	9:32:06	0	0	0	1	0	0	1
644	15分03秒	9:32:07	0	0	0	1	0	0	1
645	15分04秒	9:32:08	0	0	1	1	0	0	1

Fig. 5. Excerpt of training data.

data with deviations to a certain degree of accuracy, concerns arose regarding a potential decrease in accuracy when expanding to classifications of three or more categories. Moreover, because the determination of final productivity operations relied on pre-established rules, handling exceptional operations proved challenging. Consequently, it is difficult to assert that these rules suffice as objective criteria for judgment.

5.2. Creation of New Framework for Productive Behavior Classification System

To address these issues, we developed a new system utilizing CNNs and LSTM networks, deep learning methods that enable automatic parameter tuning through back-propagation. This approach aims to ensure objectivity and mathematical optimization by utilizing a CNN capable of retaining time-series information, thus maintaining the benefits of discrimination provided by combining mathematical methods. The system was implemented in Python, with PyTorch serving as the deep learning library of choice.

The system's design is a hybrid of CNN and LSTM, with the final output derived by integrating the outputs from each algorithm using an affine transformation. Given the lack of explicit guidance in previous studies regarding the application points of CNN + LSTM and the confidentiality surrounding the details of models involving each algorithm, we opted for a dual-algorithm approach and incorporated the results into the final output. This approach also allows considering additional methods, such as scaling the convolutional and pooled data across time windows and subsequently deriving outputs for larger window widths using LSTM.

5.3. System Development and Algorithm Selection

5.3.1. Development of Productivity Behavior Classification System

To address the previously mentioned challenges and to meet the need for a system that is both easily updatable and capable of integrating additional data types beyond acceleration, such as location information, we developed a new framework. This framework is designed to accumulate data and enable highly accurate classification over the long term. Figure 6 presents an overview of the proposed system.

Our research focuses on simple motion classification, as depicted in Fig. 6, with the ultimate objective being the development of multiple similar models. These models utilize the outputs of the five simple motions we identified and further refine their accuracy through stacking, an ensemble learning technique. Ensemble learning typically aligns the outputs of the meta-model with those of the base models. However, our system was experimentally designed to generate a broader range of outputs (three productivity behavior categories) compared to the five

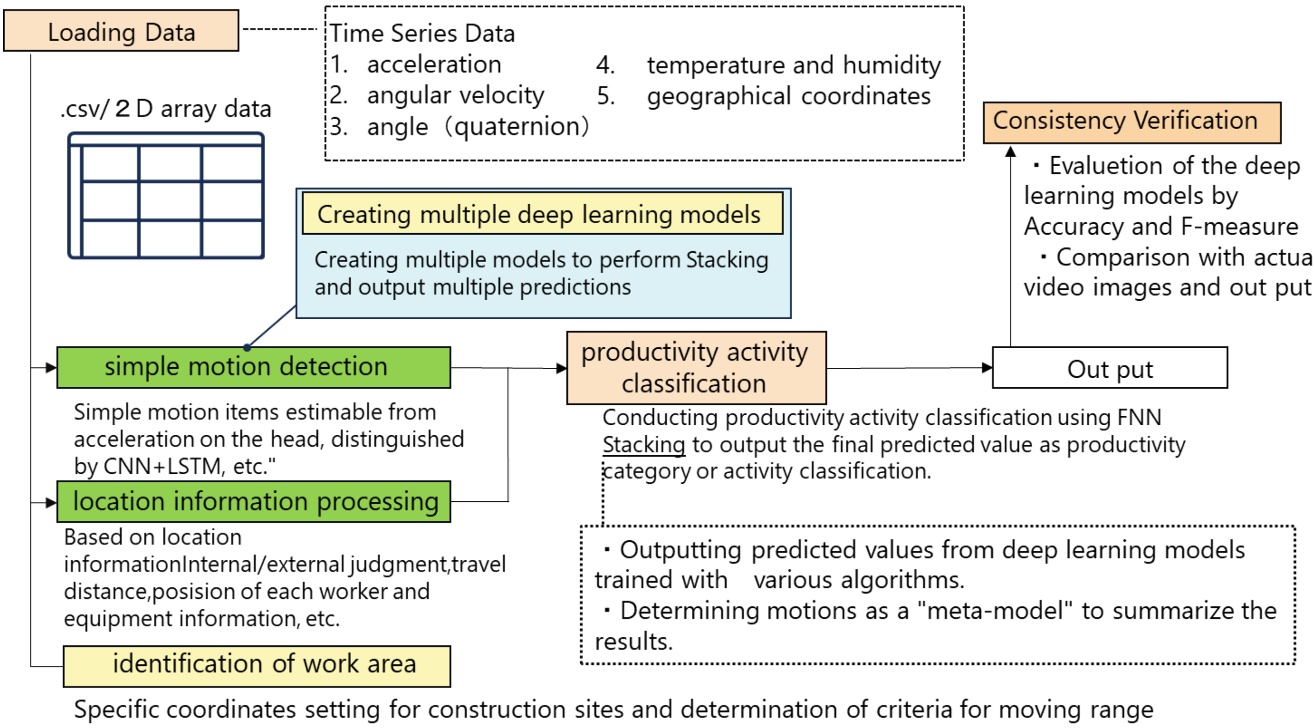

Fig. 6. System overview.

simple behavior categories identified by the base model. Additionally, this system is structured to incorporate supplementary data, such as location information, enhancing the differentiation from the base model's output.

5.3.2. Development of Productive Motion Classification System

For the classification of simple motions within the domain of HAR, we adopted CNNs and LSTM networks, both of which are deep learning methodologies previously applied in HAR research. CNNs, recognized for their efficacy in image recognition, also demonstrate significant potential in HAR by extracting features from three-dimensional motion data. In this study, we adapted CNNs, typically used for motion recognition in video footage, to analyze acceleration data. This adaptation involved treating acceleration data as image-like two-dimensional data, thereby utilizing CNNs' strengths in capturing temporal motion sequences. Our approach aimed to fulfill three key objectives: preserving essential temporal information, maintaining accurate time-series and order, and minimizing computational load by efficiently summarizing extensive temporal data.

The procedure for generating images from acceleration data is illustrated in Fig. 7. Given the continuous nature of acceleration, the data captured from the IMU in this study was sampled at 50 Hz, a limitation imposed by the sampling rate. While this discrete data is sufficiently smooth for general analysis, it does not provide sufficient features for distinguishing complex construction site activities based on productivity behavior categories. To address this, we generated images that encapsulate both waveform features and image features, enhancing feature extraction across extensive time series. This was achieved by graphing the time series of acceleration data across the X, Y, and Z axes to capture waveform features. The data transformation process begins at the first time point of the analysis window (t_1) and concludes at the last time point (t_n), with data points staggered by a shift of $\Delta t = 1$ cell, extending to t_n rows. This creates a matrix from t_1 to t_n, with time information extending to t_{2n} encapsulated in a single image. In this representation, each cell of time information progresses from the upper left corner to the lower right, with the amount of information peaking at t_n and then diminishing beyond t_{2n}. Furthermore, the visualization employs shading to reflect acceleration magnitudes, with diagonal ridges indicating waveform features. This approach not only preserves waveform characteristics but also enhances the convolutional process by emphasizing contrast variations indicative of acceleration changes, a technique known as edge processing, thereby facilitating more nuanced feature detection.

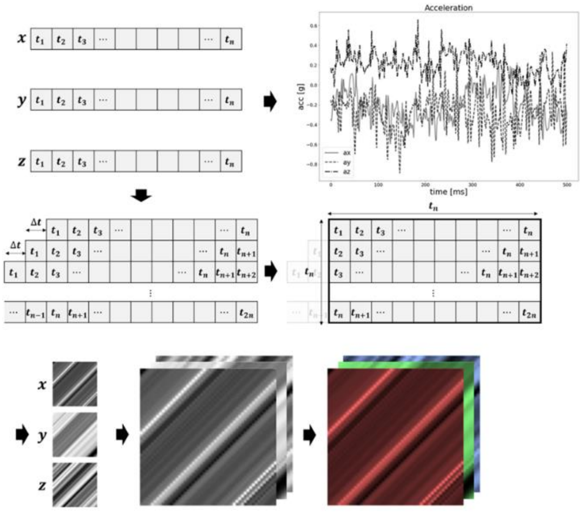

Fig. 7. Steps to generate images from acceleration data.

In this extended analysis, traditional feature extraction from simple graphs typically necessitates manual threshold settings and relies on analogies drawn from axis correlations. In contrast, our approach enables the summarization of axis-specific information through imaging techniques. We assigned the X, Y, and Z axes to the color channels red, green, and blue, respectively. This method allows for the independent representation of acceleration magnitude on each axis. Furthermore, the use of CNNs permits automatic parameter adjustments via back-propagation, significantly reducing the potential for human error to compromise the reliability of the analysis.

For preprocessing, acceleration data were segmented into 150 data point windows (equivalent to 3 s), then transformed into 150×150×3 tensors. This was followed by convolution to accentuate features, and max pooling was employed to preserve time-series characteristics. By setting the stride width to 50 data points (or 1 s), we ensured that LSTM analysis could proceed with 3 to 5 s of temporal information after imaging, thereby maintaining crucial time-series data throughout motion analysis.

To achieve generality in our model, we incorporated the rectified linear unitand the Adam optimization algorithm—both well-established in prior research—for the activation function, setting the output layer to suit specific classification needs. The model also utilized a dropout strategy in the intermediate layer, which involves randomly omitting neurons to prevent certain signals from propagating, thereby mitigating the risk of overfitting.

As briefly introduced in Section 3, LSTMs are particularly adept at processing time-series data due to their ability to retain information over extended periods. We devised a regression network that considers data from several seconds before and after the target window, incorporating LSTM to utilize its long-term data retention capabilities. This makes the LSTM model especially valuable for time-series analysis. The source code for the implementation is provided at the end of this report.

The synergy of CNN and LSTM models in our study highlights their respective strengths: CNN excels in extracting multidimensional temporal and spatial features, whereas LSTM excels in recognizing long-term dependencies within time-series data.

5.4. Simple Motion Classification Using Deep Learning

Initially, we developed a CNN-based model to distinguish between walking and working activities as a preliminary test. The training dataset comprised approximately 90,000 data points, reflecting about 30 min of activities performed by six students in our lab, all equipped similarly to field workers. This initial step aimed to ascertain the feasibility of using a single IMU dataset, fixing to helmets, for accurately classifying construction worker movements via CNN.

The results of this preliminary test are illustrated in Fig. 8. The intensity of the color corresponds to the score of the cell in each category: active work with walking, active work without walking, solely walking, and neither (indicative of minimal activity). This test achieved an overall accuracy of 87.5%, demonstrating that HAR using CNN-processed, imaged acceleration data holds promise for effective motion classification.

For the subsequent classification of five basic movements, we initially conducted a 10-fold cross-validation on the acceleration data, given the use of actual field data. This validation indicated data bias, leading us to focus solely on forward-leaning posture discrimination and walking vs. nonwalking discrimination due to the availability of ample validation data (approximately 10,000 points). The categories of repetitive motion, nonperiodic characteristic motion, and ineffective motion (which, while distinct, do not contribute to classification) were not pursued further due to insufficient data for reliable validation.

The dataset size proved inadequate for providing a robust training set. Although a basic learning experiment was conducted for these three types of movement, the majority of predictions and correct classifications were identified as true negatives (TN), indicating a general absence of the targeted activities. Consequently, the limited dataset size did not allow for achieving a high level of accuracy.

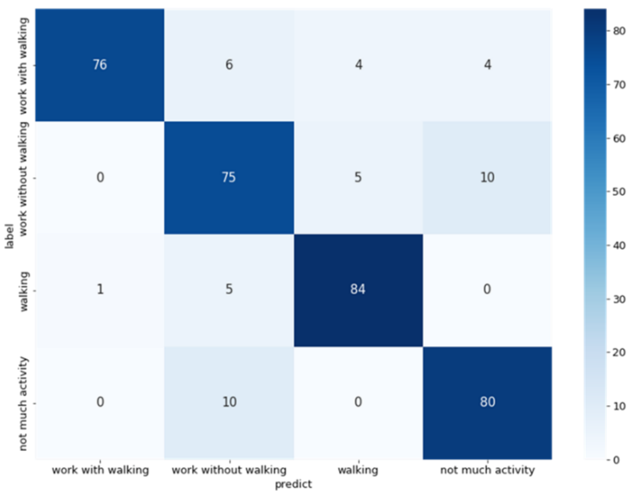

Fig. 8. Result of simple motion classification.

The evaluation of the model for classifying forward-leaning posture yielded the results depicted in Fig. 9. The model achieved a correct response rate of 92.1%, a precision rate of 98.5%, and a recall rate of 93.4%. Fig. 10 illustrates the learning curve of this model, highlighting a decreasing trend in training error alongside an increase in validation error. This pattern suggests a propensity for overfitting and a minor decline in the model's ability to generalize to unseen data.

Similarly, the model's performance in gait discrimination, as shown in Figs. 11 and 12, was commendably high, with a correct response rate of 80.0%, a precision rate of 82.2%, and a recall rate of 84.2%. However, similar to the forward-leaning posture classification, the increasing validation error across successive epochs indicates a potential reduction in generalization performance.

The methodology employed in this research did not vary across tests, with learning alternating between walking and forward-leaning based on the same acceleration data. Consequently, the learning curves for both classifications exhibited similar trends. Given the model's simplicity, it is crucial to address overfitting and enhance its capacity to generalize. This area has been identified as requiring further refinement to improve overall model performance.

5.5. Productivity Behavior Classification Using Deep Learning

The previous section highlighted challenges in acquiring sufficient data for simple motion classification. To address this, the classification of productivity behavior and work components was performed using a CNN+LSTM model on the available validation data. Figures 13–16 present the classification results and learning curves. The model achieved a correct response rate of 73.3% in distinguishing productivity behaviors, demonstrating a notable level of accuracy, particularly for supportive actions, with a recall rate of 81.5% and a precision rate of 84.6%.

Furthermore, the classification extended to specific work component categories, including "0. direct work," "1. verification of drawings and instructions," "2. moving worker movement," "3. transport of materials and equipment," and "6. waiting/standby" which correspond to the numbering shown in Table 1, for analysis. Note that only the five actions observed in actual field work were picked up, so the actions defined in Table 1 that were not observed are not shown.

The analysis identified a high verification error attributable to overfitting, a recurring issue possibly exacerbated by the model's complexity relative to the available data. CNNs inherently incorporate down-sampling to manage large datasets by reducing data volume through convolution and pooling. For instance, standard high-definition images (1920×1080) undergo convolution and pooling at intervals of 100px to manage data efficiently. In contrast, this study's generated images

DOI:10.4186/ej.2024.28.10.59

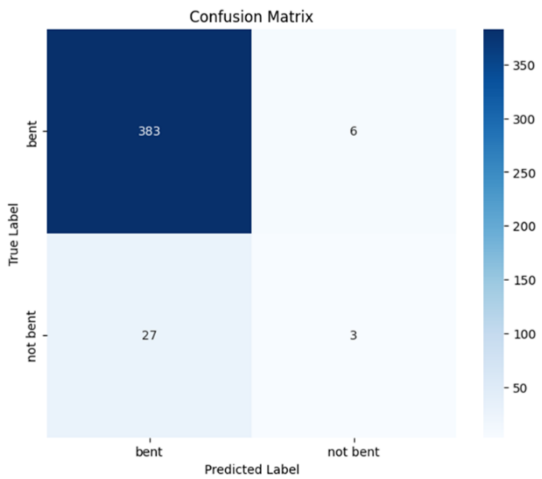

Fig. 9. Confusion matrix for forward-leaning posture classification (classification result).

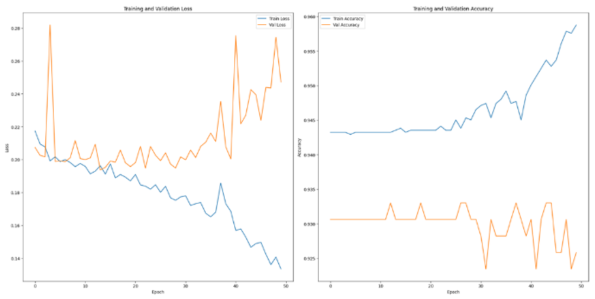

Fig. 10. Learning curve for forward-leaning posture classification model.

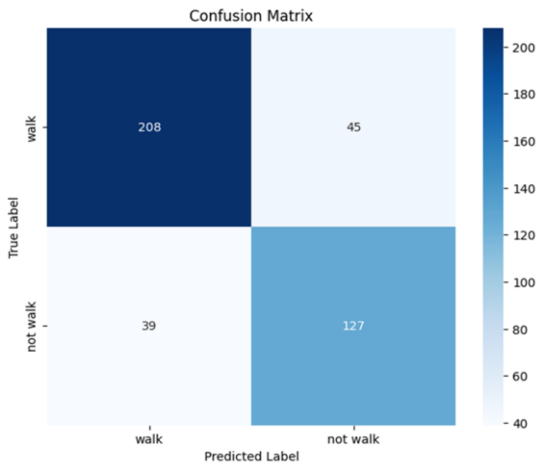

Fig. 11. Confusion matrix for walking/nonwalking classification.

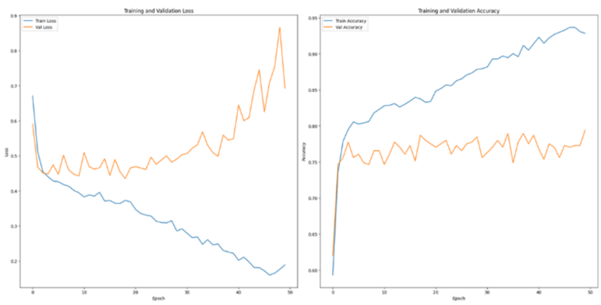

Fig. 12. Learning curves for walking/nonwalking classification model.

were 150×150 in size, emphasizing window width but potentially leading to over-extraction of features during convolution.

To mitigate overfitting, strategies such as increasing the dataset size, simplifying the CNN model, applying weight regularization, and incorporating dropout layers are recommended. Expanding the image generation beyond the 3–5 s window and adjusting the convolution kernel size could also be beneficial, although with caution, to avoid significant alterations in the window width for productivity behavior classification, which might unduly influence the model based on adjacent time-series data. Consequently, conducting control experiments to finetune the hyperparameters associated with model generation is crucial for enhancing classification accuracy and model robustness.

5.6. Discussion of Classification Test Results

The results of the classification tests demonstrate that utilizing acceleration imaging and deep learning can attain comparable levels of discrimination to those of existing systems while preserving their benefits. Although it was not possible to verify all simple motion classifications, a high level of accuracy was achieved. However, the findings also highlight the necessity for measures to mitigate overfitting. The accuracy of classifications within the productivity behavior and work component categories did not surpass that of the current system. As anticipated, the tasks executed by workers are intricate and multifaceted, with indistinct feature boundaries for discrimination, complicating effective multi-class classification.

The high accuracy in identifying forward-leaning postures is promising, yet concerns arise regarding potential variability influenced by construction site environments, work types, and occupational roles. The neck tilt was also examined as part of the forward-leaning posture classification, defined to align with data from accelerometers mounted on helmets. However, in practical scenarios, the workers' perspective varies with changes in eye level at the construction site, leading to frequent detection of neck tilt.

The analysis revealed significant variations in the manifestation of forward-leaning postures among workers engaged in pavement construction.

Figure 17 depicts a typical scene at a construction site, illustrating the variability in posture due to work demands. Operators of construction machinery often lean forward to monitor controls because these panels are positioned below eye level. The machinery's elevation relative to the pavement necessitates this forward posture for operators to effectively oversee the road surface.

Ground-level workers exhibit forward leaning for various reasons, including coordination with machinery operators positioned above them. This adjustment in posture is also evident during tasks such as shoveling, where there is considerable variance in movement. Some workers exhibit significant changes in neck and body

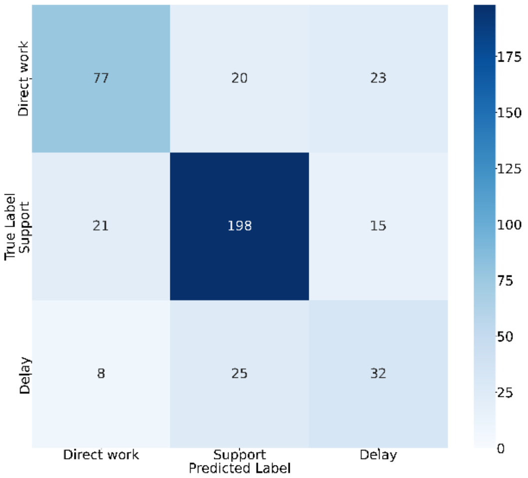

Fig. 13. Confusion matrix for productivity behavior classification model.

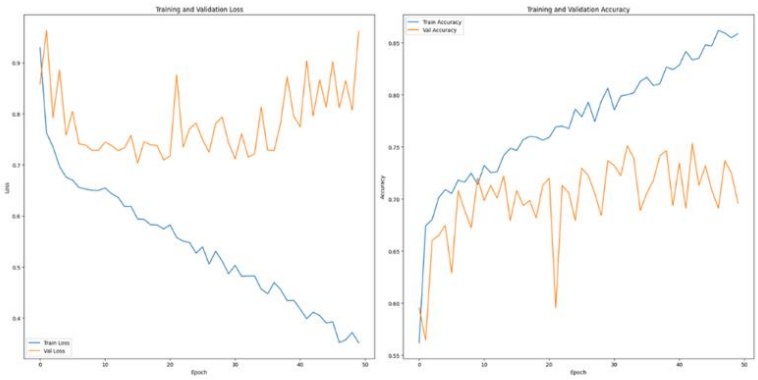

Fig. 14. Learning curve for productivity behavior classification model.

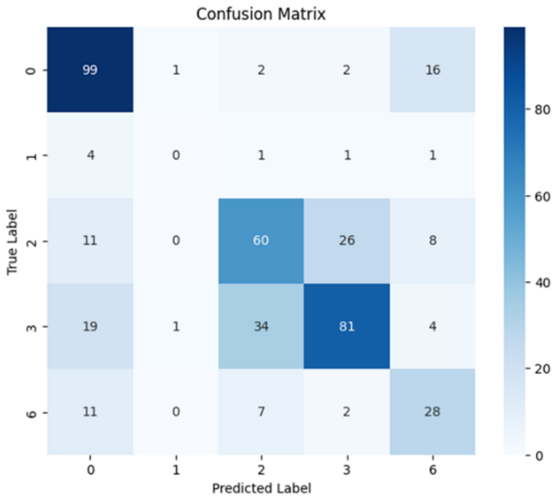

Fig. 15. Confusion matrix for classification of five work component categories.

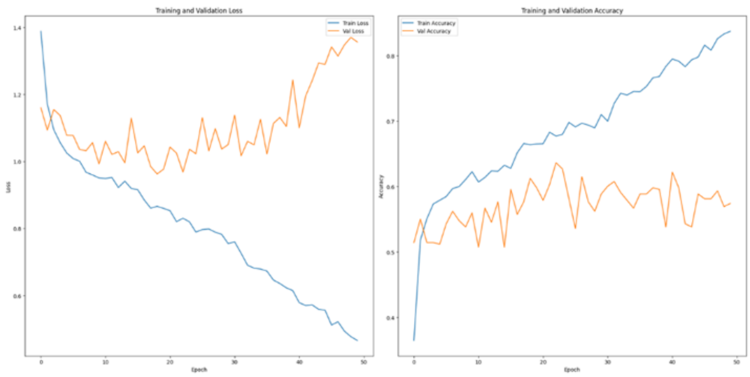

Fig. 16. Learning curve for work component category classification model.

DOI:10.4186/ej.2024.28.10.59

posture, while others maintain a consistent viewpoint with minimal posture change, focusing on the paving material.

Interestingly, while many workers adopt a neutral posture during breaks, a forward lean or downward gaze is common during meetings or when reviewing work instructions.

Previous studies using the HAR method, which relies on motion-based sensors placed on the chest, thighs, upper arms, and wrists, suggest that mounting sensors on a helmet, as done in this study, may not yield high accuracy. It is important to validate sensor data from locations other than the helmet without hindering worker mobility.

Despite these challenges, the study demonstrates that simple motion classification is feasible using helmet-mounted acceleration data alone. This finding suggests a potential for widespread acceptance of such devices in the field, given their nonintrusive nature and minimal impact on worker activity.

Fig. 17. Eye-level changing with the position of the worker and the resulting forward tilt (paving site in Japan, photo by the authors).

5.7. Evaluation of Algorithms Used and Training Data

The CNN+LSTM model utilized in this research effectively maintains time-series information, showing potential for enhancement with additional data. The CNN implementation facilitated the accumulation and improvement of training data, a notable advancement over previous systems while minimizing manual intervention. The model achieved accuracy comparable to existing systems in distinguishing between walking and forward-leaning postures, highlighting the importance of mitigating overfitting.

However, the model's performance in discriminating productivity motion categories and work components was marginally inferior to that achieved using traditional statistical methods. This discrepancy arises from the granularity of the discrimination; while prior studies classified productivity actions on a per-minute basis, our system achieved a finer resolution of 3 s, allowing for detailed discrimination at a comparable accuracy level.

Significantly, the processing time for discrimination tasks was drastically reduced: from 3 h for an hour's worth of data in previous systems to 25 min, marking a 96% reduction. This efficiency gain is attributed to the model's ability to retain time-series information while reducing the total data volume by converting acceleration data into a tensor representation of a 3 s timeframe. This approach reduces the number of data points from 450 (based on 150 data points across three axes) to a single data set, thereby decreasing the computational load.

Nonetheless, the algorithm's design, based on processing extensive information inputs, may lead to overly complex models. To enhance objectivity and accuracy, it is imperative to incorporate more objective indices into the training data and to continually collect valid data on repetitive motion, characteristic motion without periodicity, and ineffective motion. These categories, while not directly contributing to classification, are crucial for improving discrimination accuracy. Currently, the data available is insufficient to evaluate the validity of five newly identified simple motions comprehensively. To prevent overfitting and refine the model's accuracy, it is essential to continue with the systematic collection and analysis of relevant data.

6. Suggestions for System Improvements

6.1. Discussion of Individual Elements (Training Data Labels, Data Acquisition Methods, Algorithms) and Suggestions for Improvement to Increase Accuracy

This study considered the classification of simple motions as a foundational approach within the HAR methodology, employing acceleration data from a single IMU. This was used to classify productivity behaviors utilizing CNNs and LSTM networks, which are deep learning algorithms. A critical area for enhancement involves the refinement of training data labels to more accurately reflect the temporal dynamics of daily work and direct labor activities. A one-second interval often proves inadequate for reliably identifying specific operations, potentially leading to erroneous classifications. A one-minute interval may encompass multiple activities; thus, subjective determination significantly influences the identification of representative actions. Therefore, identifying an optimal time range for labeling is crucial.

Furthermore, the traditional approach of generating training data by replicating a set number of actions may not accurately capture the complexities of construction site operations. Continuous and systematic data collection from actual site workers is essential for developing a more realistic and applicable dataset.

The algorithms selected for this research, specifically CNN and LSTM, were chosen for their proficiency in processing time-series data. Other algorithms capable of handling time-series information, such as GBDT, also hold promise for effective activity discrimination. While linear models are generally favored for their reduced computational demands and enhanced processing speed, algorithms such as SVM offer high precision in binary classification at the cost of increased computational resources and processing time. GBDT and transformers represent a middle ground, offering faster processing times compared to CNN and other similar methods. Although real-time productivity analysis was not a primary focus of this study, computational efficiency and accuracy become crucial if real-time application is required.

6.2. Scope for Future

This study primarily aimed to evaluate whether productivity behavior can be accurately distinguished through the integration of an acceleration imaging process and deep learning algorithms while also addressing the limitations of existing systems. Reassessing the design of the productivity classification system, as depicted in Fig. 6, is necessary. Moreover, enhancing the precision of simple motion classification and developing a system with robust generalization capabilities requires the extraction of a broad spectrum of on-site work characteristics through ongoing data collection, correlating these characteristics with objective indicators.

One notable site characteristic observed is the collaboration within work groups. Construction site operations fundamentally involve teams composed of specific construction equipment and a designated number of workers. Thus, there exists potential to more accurately measure overall site productivity by considering not only individual worker activities but also the dynamics between workers and equipment. Incorporating such interaction data could benefit from employing location information, particularly outdoors, through methods such as GNSS or Bluetooth for inter-device communication among smartphones utilized for logging. Beyond acceleration, the IMU can detect worker proximity.

The IMU's capabilities extend to measuring magnetism and angles, suggesting that utilizing these additional data types could further refine accuracy. Specifically, analyzing posture angles, including forward-leaning positions, presents an opportunity to explore the correlation of observational changes with various factors, such as the worker's role. This goes beyond a simple binary classification of posture, inviting a more nuanced analysis.

Expanding data collection methods beyond a single accelerometer—such as incorporating sensors placed on footwear or chest-mounted accelerometers, as previously explored in sports science—could enrich the dataset and minimize erroneous classifications without overburdening workers. The system's design accommodates algorithmic updates and the inclusion of new data types, signifying an ongoing development process aimed at achieving a comprehensive system framework through continual refinement of its components.

7. Conclusion

In Japan, data on the working environment and productivity at construction sites have not yet been integrated. Productivity management data are individually collected and analyzed by on-site engineers, primarily based on attendance and piece-work records. Despite appearing fragmented, these datasets are intrinsically linked, yet their interconnections are not fully understood. This study aims to develop a system that leverages information and communication technology to automatically collect and analyze data on the working environment and productivity of construction projects. The goal is to elucidate actual working conditions at construction sites and to identify strategies for enhancing the work environment and productivity of technicians and engineers.

The study introduces the HAR method, which utilizes acceleration data from a single IMU to classify productivity behaviors with CNNs and LSTM networks. By processing acceleration data in an image-like format, the deep learning approach achieves high accuracy in classifying walking and forward-leaning postures: 92.1% accuracy, 98.5% precision, and 93.4% recall for forward-leaning postures and 80.0% accuracy, 82.2% precision, and 84.2% recall for walking. These results highlight the method's efficacy in distinguishing gait with generally high reliability. However, challenges such as mitigating overfitting and enhancing training data emerged.

To bridge the gap between existing productivity behavior categories and behaviors easily discernible through deep learning, the study introduced new behavior definitions, including walk, bent, repetitive/regular motion, characteristic motion without periodicity, and ineffective motion (significant motions not contributing to classification). Enhancing the model with these additional classifications is expected to improve the accuracy of productivity behavior categorization, previously limited by raw data analysis alone.

Given the operational constraints of construction sites, adapting the HAR method for maximal practicality necessitated deviations from conditions validated in prior research. Specifically, the placement of the IMU—on the thigh, chest, or wrist, where movements are recognized with high accuracy—was optimized for accuracy. Consequently, the deployment was restricted to a single sensor per worker.

In scenarios where individuals engage in activities voluntarily, such as sports, or for clear personal benefits, such as medical treatments, there is typically no resistance to wearing various sensors on the body. However, workers might express resistance towards having multiple sensors attached to their limbs and chests for data collection during work activities directed by an employer, as examined in this study. To mitigate this, we requested workers to wear sensors only on their helmets, presuming future integration into mandatory safety helmets.

Despite these constraints, a certain level of accuracy was achieved, suggesting that future enhancements in

DOI:10.4186/ej.2024.28.10.59

sensor placement and quantity—coupled with advancements in wearable technology and shifts in societal attitudes towards sensor acceptance—may lead to improved accuracy. Societal norms have evolved; for instance, the initial resistance to surveillance cameras in public spaces and GPS tracking has diminished over time, becoming more socially accepted for public safety and evidence gathering by the 2020s. As societal norms continue to evolve to favor the use of multiple sensors, further accuracy enhancements in data collection can be anticipated.

In the construction industry, monitoring sites through video has been challenging due to the constant changes in the construction landscape. However, for static environments such as interior construction sites, there is potential to explore HAR methods that rely on computer vision in addition to the kinematic-based methods utilized in this research.

Ultimately, the goal is to develop an integrated system capable of quantitatively and automatically monitoring the status of construction site workers to enhance both productivity and safety. As technology and devices evolve, selecting the most appropriate methods for data collection and analysis will be crucial to achieving this vision.

Acknowledgment

The author would like to thank Toshihiro Yamada and Kazunobu Mutou of Taisei Rotec Corporation for arranging the construction site for data sampling and for their advice on the analysis of pavement-related behavior. The author would also like to thank Mitsuru Hamano and Takahiro Yamaguchi of TOACK Corporation for providing the devices and technical advice. This work was supported by JSPS KAKENHI Grant Number JP21K04218 and Construction Research Institute Grant Number 2021-1.

References

[1] S. Kusayanagi, "A study on productivity improvement program for international construction projects," *Japanese Journal of JSCE*, no. 528/VI-29, pp. 143-154, 1995.

[2] Y. Gong, K. Yang, J. O. Seo, and J. G. Lee, "Wearable acceleration-based action recognition for long-term and continuous activity analysis in construction site," *Journal of Building Engineering*, vol. 52, 2022.

[3] M. Kusumoto, A. Aini, and C. Pang-Jo, "Case study of artificial intelligence at construction sites," *Artificial Intelligence and Data Science*, vol. 1, no. J1, pp. 301-306, Nov. 2020.

[4] Takenaka Corporation, "System to track the location of elevating vehicles and workers in the construction site," Available: https://www.takenaka.co.jp/news/2016/02/01/index.html. Accessed: May 10, 2022.

[5] T. Goso, A. Ochi, and S. Kusayanagi, *Journal of Infrastructure Planning and Management*, vol. 66, no. 1, pp. 317-328, 2010.

[6] T. Goso and T. Hiromitsu, "Productivity management data collection and analysis system using ZigBee and acceleration sensors at construction sites," *Monthly Automatic Recognition*, vol. 8, pp. 25-30, 2015.

[7] T. Goso, "Development and improvement of labor productivity assessment system in construction site by using information technology," *Journal of Society for Social Management Systems*, pp. SMS11-9178, 2012.

[8] S. Kusayanagi and T. Goso, "Visualization in the construction management field—Visualizing the project execution process," JSCE Magazine, Civil Engineering, vol. 96, pp. 30-32, 2011.

[9] T. Osawa and T. Goso, "Development of an automated program for work identification at construction sites," in Proceedings of the 75th Annual Conference of the Japan Society of Civil Engineers, 2020.

[10] T. Osawa and T. Goso, "Construction of a productivity classification system based on a combination of pre-classification and machine learning," in Proceedings of the Symposium on Civil Engineering Informatics, Sept. 2021, vol. 46, pp. 1-4.

[11] T. Goso, K. Muto, M. Hamano, T. Osawa, and R. Kasai, "Development of a system for collecting and analyzing labor environment and productivity data from construction sites using machine learning," *Journal of Pavement Technology*, vol. 57, no. 6, pp. 3-8, 2022.

[12] N. Hashimoto and S. Takamaeda, "Investigation of approximate computation methods for machine learning-based video processing," *IEICE Tech. Rep.*, vol. CPSY2021-59, no. DC2021-93, 2022.

[13] M. H. Uddin, J. M. Kanon Ara, M. H. Rahman, and S. H. Yang, "A study of real-time physical activity recognition from motion sensors via smartphone using deep neural network," in *Proc. 5th International Conference on Electrical Information and Communication Technology (EICT)*, 2021.

[14] K. Lee and S. Han, "Convolutional neural network modeling strategy for fall-related motion recognition using acceleration features of a scaffolding structure," *Automation in Construction*, vol. 130, Oct. 2021.

[15] R. Imai, D. Kamiya, H. Inoue, S. Tanaka, J. Sakurai, K. Sakamoto, T. Fujii, K. Mimura, and M. Ito, "Study on trial of person identification for safety management at construction sites," in Proceedings of the Symposium on Civil Engineering Informatics, 2018, vol. 43, no. 37, pp. 145-148.

[16] R. Sanford, S. Gorji, L. G. Hafemann, B. Pourbabaee, and M. Javan, "Group activity detection from trajectory and video data in soccer," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020.

[17] Y. Jin, G. Suzuki, and H. Shioya, "Detecting and visualizing stops in dance training by neural network

based on velocity and acceleration," *Sensors*, vol. 22, no. 5402, 2022.

[18] M. A. Khatun, M. A. Yousuf, S. Ahmed, M. Z. Uddin, S. A. Alyami, S. Al-Ashhab, H. F. Akhdar, A. Khan, A. Azad, and M. A. Moni, "Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, 2022.

[19] S. Joshi and E. Abdelfattah, "Deep neural networks for time series classification in human activity recognition," in *IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2021, pp. 559-566.

[20] W. Zhai, "Human activity recognition based on acceleration sensor and neural network," in *8th International Conference on Orange Technology (ICOT)*, 2020.

[21] M. Kandpal, R. K. Barik, S. S. Patra, B. Sharma, S. Chowdhury, and I. B. Dhaou, "Human activity recognition in smart cities from smart watch data using LSTM recurrent neural networks," in *1st International Conference on Advanced Innovations in Smart Cities (ICAISC)*, 2023.

[22] C. O. Odhiambo, S. Saha, C. K. Martin, and H. Valafar, "Human activity recognition on time series accelerometer sensor data using LSTM recurrent neural networks," in *CSCE Conference Spring Nature*, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2206.07654

[23] S. W. Pienaar and R. Malekian, "Human activity recognition using LSTM-RNN deep neural network architecture," in *IEEE 2nd Wireless Africa Conference (WAC)*, 2019.

[24] Y. Takahashi, K. Nakamura, T. Kamiyama, M. Oguchi, and S. Yamaguchi, "Person identification based on accelerations sensed in smartphones with LSTM," *Journal of Information Processing*, vol. 29, pp. 707–716, Oct. 2021.

[25] S. S. Vakacherla, P. Kantharaju, M. Mevada, and M. Kim, "Single accelerometer to recognize human activities using neural networks," *Journal of Biomechanical Engineering*, vol. 145, no. 6, p. 061005, Jun. 2023. doi: 10.1115/1.4056767.

[26] T. Hayama and R. Arakawa, "Walking-posture classification from single-acceleration sensor data using deep learning and its evaluation," *International Journal of Service and Knowledge Management*, vol. 5, no. 2, pp. 23-36, 2021.

[27] T. Gondo and R. Miura, "Accelerometer-based activity recognition of workers at construction sites," *Frontiers in Built Environment*, vol. 6, p. 563353, 2020. doi: 10.3389/fbuil.2020.563353.

[28] K. Oi, M. Matsushita, H. Hori, and Y. Takaoka, "Detecting fall motion from gait data using a wearable sensor," in *Proceedings of the 28th Interactive Information Access and Visual Mining Research Conference*, The Japanese Society for Artificial Intelligence, 2022, pp. 1-8.

[29] T. Kawamura, S. Tsuchida, T. Terada, and M. Tsukamoto, "Verification of equipment wearing bias in motion measurement experiments using wearable sensors," IPSJ SIG Technical Report, pp. 21–30, 2021.

[30] K. Kamoto and M. Arai, "Construction machinery active safety technology—Development of machine monitoring system," AKENAKA Technical Research Report, no. 67, 2011.

[31] K. Kitamura, T. Odaka, J. Kuroiwa, I. Suwa, and H. Shirai, "A method to human activity recognition using acceleration data by machine learning," Memorandum of Graduate Engineering, University of Fukui, vol. 68, pp. 59-65, 2020.

[32] S. Aihara, K. Ishibe, R. Sabu, and H. Iwata, "Deep learning method for karate motion identification using inertial sensor data," *Transactions of the JSME*, vol. 78, no. 903, 2021.

[33] E. Nakahara, T. Shioda, and M. Toyoda, "A study of explainable anomaly detection for multivariate time series," in *Forum on Information Technology 2021 (FIT)*, The Institute of Electronics, Information and Communication Engineers, vol. 2, pp. 423-429.

[34] K. Zhang, T. Umezawa, and N. Osawa, "An evaluation of the influence of dimensionality of features and time-series prediction models in human activity recognition," in *Forum on Information Technology 2019 (FIT)*, The Institute of Electronics, Information and Communication Engineers, vol. 2, pp. 287-288.

[35] N. Hirasawa and D. Shimizu, "Visualization of dancer's emotional state and cultural support using deep learning in breakdance," IPSJ SIG Technical Report, vol. 2018-HCI-179, no. 4, 2018.

[36] F. Hajjej, M. Javeed, A. Ksibi, M. Alarfaj, K. Alnowaiser, A. Jalal, N. Alsufyani, M. Shorfuzzaman, and J. Park, "Deep human motion detection and multi-features analysis for smart healthcare learning tools," *IEEE Access*, vol. 10, 2022.

[37] S. Agatsuma, M. Nakagawa, T. Ono, S. Takahashi, B. Shizuki, and S. Saga, "Classification of rubbing objects using wireless accelerometers and convolutional neural network," in *ROBOMECH 2018 Collection of Lecture Abstracts*, 1P1-K14-, 2018.

[38] T. Nishida, K. Seeto, N. Yabuki, and T. Hujishima, "Visualization of construction sites using video recognition AI for productivity support systems," in *Proceedings of the Symposium on Civil Engineering Informatics*, vol. 44, pp. 177-180, 2019.

[39] R. Tanaka, N. Yabuki, and T. Fukuda, "Development of a workers' posture estimation system based on the OWAS method using helmet-mounted terminal," *JSCE Transactions F3 (Civil Engineering Informatics)*, vol. 77, no. 2, pp. 114-123, 2021.

[40] H. Kasahara, T. Ito, H. Yoshii, K. Tsurumaru, and N. Komatsu, "A study on the walking states classification method using cycles of acceleration

DOI:10.4186/ej.2024.28.10.59

data," *IEICE Technical Report*, vol. 111, no. 50, LOIS2011-3, pp. 21-26, May 2011.
[41] K. Murao and T. Terada, "A motion recognition method by constancy decision," *IPSJ Journal*, vol. 52, no. 6, pp. 1968-1979, 2011.
[42] L. Liao, D. Fox, and H. Kautz, "Location-based activity recognition," in *Advances in Neural Information Processing Systems*, 2005.
[43] H. Chena, S. H. Cha, and T. W. Kim, "A framework for group activity detection and recognition using smartphone sensors and beacons," *Building and Environment*, vol. 158, pp. 205-216, 2019.
[44] C. Sabillon, A. Rashidi, B. Samanta, M. A. Davenport, and D. V. Anderson, "Audio-based Bayesian model for productivity estimation of cyclic construction activities," *Journal of Computing in Civil Engineering*, vol. 34, no. 1, 2020.
[45] C.-F. Cheng, A. Rashidi, M. A. Davenport, and D. V. Anderson, "Activity analysis of construction equipment using audio signals and support vector machines," *Automation in Construction*, vol. 81, pp. 240-253, 2017.
[46] L. Joshua and K. Varghese, "Construction activity classification using accelerometers," in *Construction Research Congress*, 2010, pp. 61-70.
[47] C. Hernandez, T. Salton, V. Balali, and R. Akhavian, "A deep learning framework for construction equipment activity analysis," *Journal of Computing in Civil Engineering*, pp. 479-486, 2019.
[48] K. M. Rashid and J. Louis, "Automated activity identification for construction equipment using motion data from articulated members," *Frontiers in Built Environment*, vol. 5, article 144, Jan. 2020.
[49] H. Kurasawa, Y. Kawahara, H. Morikawa, and T. Aoyama, "User posture and movement estimation based on 3-axis acceleration sensor position on the user's body," *IPSJ SIG Technical Report*, vol. 2006-UBI-11, no. 3, 2006.
[50] R. Sato, H. Habe, I. Mitsugami, S. Satake, K. Sumi, and Y. Yagi, "Social group discovery extracting useful features using multiple instance learning," Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, vol. 28, no. 6, pp. 920-931, 2016.

Ryuji Kasai, photograph and biography were not available at the time of publication.

Takashi Goso, photograph and biography were not available at the time of publication.

Mizuki Akiyama, photograph and biography were not available at the time of publication.

Hiroto Kato, photograph and biography were not available at the time of publication.