Corresponding author:
Chiuhsiang Joe Lin;
Email: cjoelin@mail.ntust.edu.tw

Machine learning models for sequential motion recognition in human activity for logistics operations efficiency improvement

Chih-Feng Cheng ![ORCID], Chiuhsiang Joe Lin ![ORCID] and Qin-Xuan Hu

Department of Industrial Management, National Taiwan University of Science and Technology, Taipei, Taiwan

Abstract

Human activity recognition (HAR) is a vital component of human–robot collaboration. Recognizing the operational elements involved in an operator's task is essential for realizing this vision, and HAR plays a key role in achieving this. However, recognizing human activity in an industrial setting differs from recognizing daily living activities. An operator's activity must be divided into fine elements to ensure efficient task completion. Despite this, there is relatively little related research in the literature. This study aims to develop machine learning models to classify the sequential movement elements of a task. To illustrate this, three logistic operations in an integrated circuit (IC) design house were studied, with participants wearing 13 inertial measurement units manufactured by XSENS to mimic the tasks. The kinematics data were collected to develop the machine learning models. The time series data preprocessing involved applying two normalization methods and three different window lengths. Eleven features were extracted from the processed data to train the classification models. Model validation was carried out using the subject-independent method, with data from three participants excluded from the training dataset. The results indicate that the developed model can efficiently classify operational elements when the operator performs the activity accurately. However, incorrect classifications occurred when the operator missed an operation or awkwardly performed the task. RGB video clips helped identify these misclassifications, which can be used by supervisors for training purposes or by industrial engineers for work improvement.

![CAMBRIDGE UNIVERSITY PRESS]

Introduction

Industry 4.0 aims to innovate industries through the Internet of Things (IoT), cyber-physical systems, artificial intelligence (AI), Big Data, cloud computing, and robotics (Iqbal et al., 2022; Maddikunta et al., 2022; Nahavandi, 2019), with a focus on enhancing fabrication efficiency via automation (Maddikunta et al., 2022; Nahavandi, 2019). This includes the implementation of robots to replace human workers in the Industry 4.0 paradigm.

Bringing human workers back to factories

However, Industry 4.0 has not yet been successfully implemented worldwide (Iqbal et al., 2022; Nahavandi, 2019). The first reason is that Industry 4.0 involves the integration of various advanced technologies, requiring support from infrastructures that are not yet available in most countries outside the United States and European Union (Chien et al., 2017). Second, the highly customized demands of modern businesses (Alves et al., 2023; Gan et al., 2023; Maddikunta et al., 2022) make fully automated systems prone to failure. Automation systems with high buildup costs cannot gain an advantage in the small production quantities of customized orders (Iqbal et al., 2022). The third reason is that Industry 4.0 does not strongly consider environmental sustainability (Nahavandi, 2019). The reduction of the carbon footprint in manufacturing is receiving greater attention from public interest groups, international organizations, and governments (Nahavandi, 2019). Consumers are also increasingly supporting environmentally friendly companies (Nahavandi, 2019). Balancing environmental protection with high-efficiency production will be a critical issue in modern industry.

Furthermore, the biggest issue may be the ethical problem of removing human workers from the production line. In the conceptual framework of Industry 4.0, humans are considered obstacles to improving productivity. As a result, automation is prioritized over human labor, leading to a decline in employment rates (Iqbal et al., 2022). This approach faces opposition from labor unions, congress, and governments (Iqbal et al., 2022; Nahavandi, 2019).

Nevertheless, practitioners have become aware of the drawbacks associated with the original concepts of Industry 4.0. Consequently, they have revised the interactions between humans and machines (Romero et al., 2016). The main modification emphasizes a human-centric approach, as proposed by a nonprofit, industry-driven association in Europe (Romero et al., 2016). Recently, Industry 5.0, which builds upon the technologies developed in Industry 4.0 (Iqbal

et al., 2022), has emphasized the vision of creating sustainable, human-centric, and resilient industries. The most prominent aspect of Industry 5.0 is the reintegration of human workers into shop floors, which has garnered significant interest from researchers (Alves et al., 2023; Cimino et al., 2023; Longo et al., 2020; Moller et al., 2022; Papetti et al., 2020; Romero and Stahre, 2020). The primary goal is to leverage human creativity and intelligence in smart manufacturing systems (Iqbal et al., 2022), thereby ensuring the well-being of workers (Moller et al., 2022).

A human-centric solution for the smart factory

A human-centric solution in Industry 5.0 is an autonomous production system with humans in and on the loop (Nahavandi, 2019). Humans and machines coexist and cooperate (Demir et al., 2019; Johri et al., 2021) based on Industry 4.0, forming a human–cyber-physical system (HCPS; Romero et al., 2016). The robots in the HCPS collaborate with human workers, acting as another worker who stands beside them to assist in finishing a job. These collaborative robots, called cobots, are equipped with human intelligence by machine learning algorithms. Cobots not only observe and learn how an operator performs a task but also have machine recognition (Maddikunta et al., 2022; Nahavandi, 2019). Cobots receive data from wearable sensors on human workers (Fortino et al., 2015; Kong et al., 2018; Maddikunta et al., 2022; Montini et al., 2022). The AI algorithm processes the data so that cobots can locate the operator's position, analyze the worker's posture, know the goal of the ongoing task, and perceive the human coworker's expectations. With this information, cobots autonomously decide what to do and how to assist human workers (Maddikunta et al., 2022; Nahavandi, 2019). Accordingly, wearable sensors and wireless communication protocols in IoT, and machine recognition algorithms in AI play essential roles in the human-centric solution in Industry 5.0.

The challenges of human activity recognition (HAR) of human–robot collaboration

HAR has been recognized as an essential component of human–robot collaboration (Lucci et al., 2022; Martinez-Villasenor and Ponce, 2019). However, HAR in the industrial sector differs from the recognition of activities of daily living (ADL), which has been well-documented in the literature (Al-Amin et al., 2020). ADL literature focuses on the recognition of a complete and unsegmented action. In industry, a human operator's activity within a task needs to be divided into fine elements so that the collaborative robot can respond to each recognized element. One approach is to apply gesture recognition in human–robot collaboration (Liu and Wang, 2018). However, gestures are subtasks that may impact the operator's performance in the main task. Additionally, requiring operators to memorize gestures may increase their mental workload (Lucci et al., 2022).

A review of HAR (Wang et al., 2019) has reported that most research has focused on people's ADL. With technological advancements, HAR applications have rapidly developed in various fields, including security and surveillance, smart homes, entertainment, medical healthcare, and gait analysis (Wang et al., 2019). Unfortunately, these types of HAR have limited impact on human–robot collaboration, as operators on a shop floor perform tasks with many continuous elements.

Moreover, most existing public databases are based on ADL (Anguita et al., 2013; Banos et al., 2014; Chavarriaga et al., 2013; Micucci et al., 2017; Reiss and Stricker, 2012; Stisen et al., 2015; Sztyler and Stuckenschmidt, 2016). In comparison to ADL, databases or research focused on manufacturing or standard operating instructions on a shop floor are relatively limited (Zappi et al., 2008).

Worker and machine process chart for HAR in human–robot collaboration

To realize the vision of human–robot collaboration, the cobot must identify and understand every element being processed by the human worker. This requires decomposing activities into various operational elements, similar to how an industrial engineer creates a worker and machine process chart (Freivalds, 2009) using motion and time study methods. With this information, the cobot can perceive which element the operator is working on and autonomously decide how to collaborate with the worker, rather than relying on preprogrammed responses.

An important task for industrial engineers is to establish standard operating procedures (SOPs) for various production plants, ensuring high-efficiency performance while considering the welfare of workers. By utilizing professional methods from motion and time study, engineers first observe the sequence of workers' actions, breaking them down and categorizing them into a series of sequential motions. The time required for each sequential motion is then measured. These sequential motions are further evaluated to determine if they can be optimized through the principles of elimination, rearrangement, combination, or simplification. Subsequently, using a selected benchmark worker, the industrial engineer uses a stopwatch to measure the time required for each sequential motion and allows for a certain margin, establishing the standard time for each task. Production site managers can use this SOP and timing as training material for new employees. For employees whose performance does not meet the standard, a comparison of their sequential motions can help identify whether their actions are correct, and comparing the time taken for each sequential motion with the standard time can guide them toward improvement.

However, classical motion and time study has some drawbacks. Industrial engineers traditionally use stopwatches and manual paper records to observe workers' actions and timing, which incurs significant labor and time costs, introduces subjective variations, and can reduce accuracy due to visual errors. The presence of engineers in the workplace may make workers uncomfortable, affecting their performance, and even creating tension or conflict, leading to a stressful work atmosphere.

This research proposes integrating the concept of motion and time study with sequential motion recognition in human activity, which not only promises to realize the human–robot collaboration envisioned in Industry 5.0 but also helps workers improve their performance, enables site managers to create more effective training programs, and assists industrial engineers in completing system improvements more efficiently.

In short, cobots must be equipped with HAR intelligence that can recognize processing elements in specific tasks to achieve effective human–robot collaboration in a smart factory.

Motivation and objective

The vision of smart factories within the Industry 5.0 framework involves applying human-centric solutions to enable autonomous human–robot collaboration. Recognizing the operational elements of a task performed by an operator using HAR methods is essential to realizing this vision. However, there is relatively little research on

this topic in the literature. Therefore, before achieving human–robot collaboration, this study is motivated by the need to fill this critical research gap.

This study addresses the limitations of traditional methods by employing machine learning models for HAR with intelligent technology. As a pilot study combining the concepts of motion and time study and HAR, this research focuses on sequential motion recognition. Once the motion recognition technology is successfully developed, the operational time for each sequential motion can be estimated based on the number of data samples and sampling frequency, thereby achieving the goal of improving production performance. If the method proposed in this study successfully achieves the goal of "sequential motion recognition in human activity," it would demonstrate that this innovative concept could be extended to any manufacturing industry that utilizes SOPs.

The aim of this research is to apply machine learning methods to recognize sequential motions in operations, aligning with the vision of human–robot collaboration in Industry 5.0. Additionally, the research seeks to assist industrial engineers in developing SOPs, training personnel, and improving performance in the workplace. HAR automates information gathering, saving time and avoiding disputes by replacing time-consuming and potentially controversial traditional motion and time study methods. Additionally, HAR results provide valuable feedback for new employee training, enabling supervisors to design programs that improve performance and productivity on the shop floor.

Related works

To develop machine learning models for HAR, data on human movement are collected using various types of devices and then fed into training models. These devices can be divided into two categories: vision-based and sensor-based. Preprocessing and feature extraction are vital for the collected data before it is used for model training. In this section, we discuss the literature related to these topics.

Data collecting devices

Vision-based sensors in HAR

In vision-based action recognition, traditional RGB cameras are popular and affordable for HAR researchers. The relevant literature has explored various features of RGB images and classification methods (Aggarwal and Ryoo, 2011; Poppe, 2010; Ramanathan et al., 2014; Weinland et al., 2011). However, intensive image processing and computer vision algorithms require considerable hardware resources, and RGB images, which lack 3D motion data, need to be processed (Aggarwal and Xia, 2014).

Compared to RGB cameras, three categories of devices providing 3D kinematics information of the human body have been widely used in HAR research. The first category includes the expensive optical motion capture (OMC) system. This method involves attaching reflective markers to specific joints on the human body and using multiple infrared cameras to estimate the 3D coordinates of each joint. The OMC method is considered the gold standard in biomechanics studies, and many HAR studies using OMC are available in the literature (Aurand et al., 2017; Zhu and Wang, 2011).

The second category of equipment utilizes stereo cameras to obtain 3D data, including depth, through stereo matching and depth calculation (Argyriou et al., 2010). However, stereoscopic 3D reconstruction algorithms are computationally intensive and

extremely sensitive to the illumination and complexity of the background environment (Aggarwal and Xia, 2014), which makes them less prevalent today.

A third approach is based on distance or depth sensors. In the last decade, low-cost RGB depth cameras have become highly cost-effective and have been used for real-time 3D motion capture in HAR studies (Jalal et al., 2017; Khan et al., 2022; Qi et al., 2022; Yadav et al., 2021). Compared to traditional RGB images obtained by conventional cameras, the depth images generated by depth cameras are less sensitive to changes in illumination, and the performance of HAR is higher since the 3D kinematics of joints are utilized (Shotton et al., 2013).

Although vision-based methods are still popular in HAR studies, they face several challenges that limit their performance. For example, obstructions between the human body and the camera, variations in movement between different individuals, and cluttered backgrounds can all hinder effectiveness (Vrigkas et al., 2015). Furthermore, vision-based methods are limited by the camera's field of view and the effective space of the camera setup. These limitations make it difficult for vision-based HAR to be applied outside the laboratory.

Sensor-based sensors in HAR

To overcome vision-based device limitations, recent years have seen the development of wearable inertial sensors equipped with accelerometers and gyroscopes. These sensors not only provide 3D acceleration and angular velocity data but also derive 3D joint coordinates, similar to depth cameras. This shift has gained traction in the HAR community, offering a notable advantage for field studies (Xu et al., 2017). Numerous studies leverage portable devices such as smartphones, smartwatches, and inertial measurement units (IMUs) with embedded inertial sensors for data collection and analysis (Baldominos et al., 2019; Burns and Whyne, 2018; Reyes-Ortiz et al., 2016; San-Segundo et al., 2016; Shoaib et al., 2016; Xu et al., 2017).

The sensor technology efficiently addresses broader areas and dynamic lighting conditions. Advancements in low power consumption and computational capabilities empower inertial sensing for long-term recording, computation, and continuous interaction with individuals. In addition to three-axis acceleration and angular velocity, wearable inertial sensors can also derive human joint 3D coordinates, similar to those obtained from depth cameras.

Data preprocessing and feature extraction

HAR involves a sequential chain of activities, beginning with capturing raw data from various sensors, followed by preprocessing to eliminate discomfort and noise, segmenting time-series data for target recognition, extracting features based on data attributes, reducing data dimensionality for feature refinement, building and training the classification model, and finally, verifying model performance (Bulling et al., 2014).

Sliding windows play a crucial role in time series data classification and are commonly used in segmentation methods. These windows can be nonoverlapping or have fixed-size overlapping structures. Determining the optimal window size is vital, considering variations in sensor placement and activity periods. Empirical determination of window length is a practical approach (Burns and Whyne, 2018; Jalal et al., 2017; Janidarmian et al., 2017).

Human activities form a time series data set, necessitating analysis through time windows rather than individual data frames.

Feature extraction is essential for comparing activity differences, with widely applied time-domain features, such as median, mean, standard deviation, entropy, kurtosis, and skewness. Frequency domain features, including spectral entropy, peak power, and peak frequency, have gained popularity. Additionally, wavelet domain features derived through wavelet transform are explored in the literature (Burns and Whyne, 2018; Hassan et al., 2018; He et al., 2018; Ignatov, 2018; Wang et al., 2017).

In summary, the literature highlights the significance of sliding windows in HAR research, discussing their use, distinctions between nonoverlapping and fixed-size overlapping windows, the critical importance of determining the optimal window size for effective data classification, and the impact of window overlap on model performance.

Machine learning models for HAR

HAR involves understanding and classifying actions based on data from human movement, typically framed as a pattern classification problem solvable by machine or deep learning models (Konar and Saha, 2018).

AI algorithms used in HAR are divided into traditional machine learning-based and deep learning-based methods. Recent review papers have surveyed the models used in the literature (Bozkurt, 2021; Dua et al., 2022; Kulsoom et al., 2022). Frequently utilized in research studies are support vector machine (SVM), k-nearest neighbor (kNN), Gaussian mixed model (GMM), decision tree (DT), naïve Bayes (NB), random forest (RF), and hidden Markov model (HMM). Moreover, developing new deep learning-based models is a hot research topic in related communities. For deep learning-based methods, researchers often implement convolutional neural networks (CNN), deep neural networks (DNN), recurrent neural networks (RNN), deep belief networks (DBN), and long short-term memory (LSTM) in their investigations. Combining two algorithms to improve the model's performance is also common. For example, LSTM is frequently embedded in CNN or RNN, and hybrid methods with SVM, HMM-SVM, or kNN-SVM are regularly adopted in machine learning models.

The field of HAR has witnessed significant advancements in recent years, driven by the proliferation of wearable sensors and the remarkable capabilities of deep learning algorithms. Deep learning has emerged as a powerful tool for extracting meaningful features from sensor data and accurately classifying human activities.

Studies on sequential motion recognition in human activity

The majority of HAR research has traditionally focused on ADL (Wang et al., 2019). However, recent studies have shifted toward decomposing and recognizing elements within individual activities. One study proposed an HMM-based classification system for online classification of IMU and force sensor data from a lower limb exoskeleton, achieving an average accuracy of 84.5% in walking motion element recognition (Beil et al., 2018).

In a study on industrial lifting operation involving three elements, acceleration measurements were captured and used to develop a classifier within the HMM framework, achieving an impressive average accuracy of 99.8% (Ishibashi and Fujii, 2021).

Another study assessed the recognition of eight elements in martial arts using motion data from IMUs. Among various sequence-based methods, the features of kinematics data extracted by a dynamic time warping (DTW) algorithm with kNN classifier achieved the highest accuracies 99.6% (Li et al., 2022).

A separate study proposed a method for recognizing a nine-element industrial assembly activity by combining CNN and LSTM techniques with IMU and video data. The model achieved accuracies of 88% for IMU-based skeleton data, 77% for RGB spatial video, and 63% for RGB video-based skeleton (Tuli et al., 2022).

In a follow-up study of Tuli et al. (2022), a method was introduced for automatically generating process time analyses of manual assembly processes with five elements. This method detected approximately 76% of manual operations in a proposed use case scenario (Jonek et al., 2023).

In contrast to recognizing entire elements, human–robot collaboration often relies on time stamp-based activity recognition, as validated by a single participant. One study introduced a general library of atomic predicates, which can be combined with first-order logic for modeling general industrial assembly processes in human–robot collaboration using RGBD data. The methodology achieved average accuracies of 99.3% in a complex collaborative assembly use case involving seven elements (Lucci et al., 2022).

Another study presented an innovative operator advice model utilizing deep learning. This model incorporated three independent cameras and a decision tree classifier as a motion recognition mechanism. Applied to a human–robot collaborative task with five manual and four robot elements, the model demonstrated promising performance with an accuracy of 95.1% across the five manual elements (Wang and Santoso, 2022).

In summary, breaking down a human operator's task into finer elements is essential for efficient analysis by industrial engineers. Within the framework of Industry 5.0, research related to the decomposition and recognition of operators' actions is particularly important.

Materials and methods

This study simulates actual semiconductor IC product shipment manual operations to validate the effectiveness of the HAR machine learning model. Participants, wearing IMUs in the laboratory, completed three tasks according to the SOP. The kinematics data collected during this process follow the activity recognition development process proposed by Bulling et al. (2014), which includes preprocessing raw data from motion capture devices, segmenting time-series data, extracting features based on data attributes, establishing and training an activity classification model, and verifying its performance.

Participants

Sixteen participants (8 males, mean height = 174.6 cm, SD = 5.3 cm; 8 females, mean height = 161.6 cm, SD = 4.7 cm), aged 18–50, underwent simulated tasks. Wearable sensors (XSENS) with 13 upper-body sensors collected acceleration, angular velocity, and quaternion data at 60 Hz (Figure 1). Supplementary spatial coordinates were obtained from OMC and Microsoft Kinect V2, placed 3 m in front of participants at a height of 1.5 m, with sampled at 30 Hz.

The movements of three laboratory members, who were trained as benchmark workers according to the SOP, serve as the training dataset for the machine learning model.
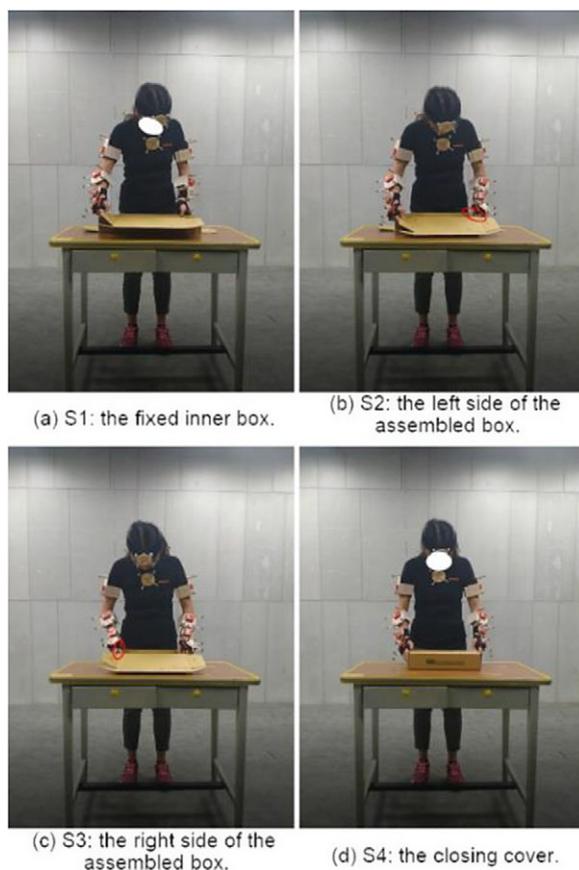
Logistic tasks

Inspired by the author's previous work experience, this study uses the shipping operations of a semiconductor design company as a
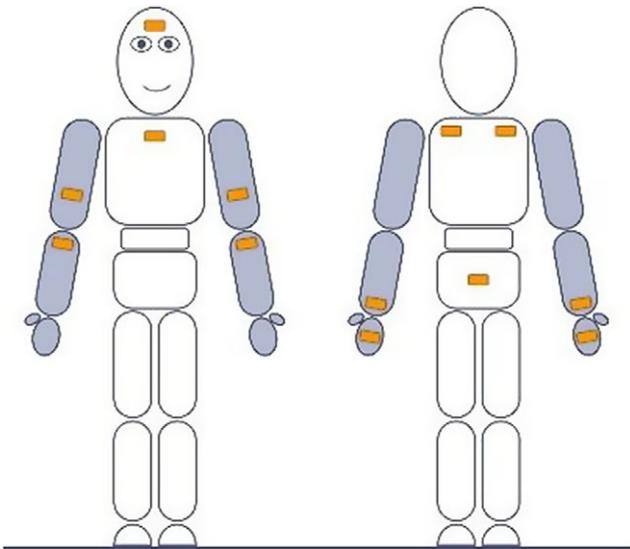
Figure 1. The positions of 13 XSENSs on the front and back of the human body.

case study to preliminarily verify the feasibility of the proposed research idea. Among the various operations involved in shipping to end customers, preparing and folding large and small cardboard boxes, placing goods inside, and sealing the boxes represent the basic building blocks of logistics operations, which are often performed manually. This study mimicked logistics operations in semiconductor design houses, specifically focusing on adherence

(a) S1: the fixed inner box.

(b) S2: the left side of the assembled box.

(c) S3: the right side of the assembled box.

(d) S4: the closing cover.

Figure 2. Operational elements decomposition of the assembly of the inner box (T1).

to the JEDEC packaging standard for shipping. The standard involves using a JEDEC-6A outer carton containing six JEDEC-6B inner boxes, each housing ICs. The research simulated tasks such as the assembly of inner boxes (T1), outer carton assembly (T2), and final carton packing for shipment (T3). Although these operations are relatively simple, they reflect real-world tasks encountered by the author during industrial work, and the boxes used are identical to those in real-world applications.

The SOP for assembly of inner boxes comprises four elements (Figure 2): Element 1 (S1) is "fixed inner box material," Element 2 (S2) is "assemble the left side of the inner box," Element 3 (S3) is "assemble the right side of the inner box," and Element 4 (S4) is "close the top cover." Outer carton assembly consists of three elements in the SOP (Figure 3): Element 1 (S1) is "fixed outer carton material," Element 2 (S2) is "sealing the carton bottom," and Element 3 (S3) is "turn the bottom of the carton facing upward." Final carton packing consists of two elements in the SOP (Figure 4): Element 1 (S1) is "placing the inner boxes into the carton," and Element 2 (S2) is "sealing the carton." Participants executed T1, T2, and T3 sequentially, following instructions from the experimental controller. They maintained a natural stance with hands hanging down for 1–2 s at the beginning and end of each task.

Data preprocessing

Research by Wang and Santoso (2022) and Hashemi (2019) suggests that maximum absolute (MA) scaler normalization yields higher recognition accuracy than min–max (MM) normalization. Therefore, this study employs both normalization methods and compares their outcomes.

In the segmentation phase, the commonly used sliding window method is employed, which involves distinguishing between non-overlapping and fixed-size overlapping sliding windows. Determining the optimal window size is crucial, with studies by Janidarmian et al. (2017) indicating that this varies based on body sensor locations. Increasing the overlap in the sliding window improves model performance, as noted by Burns and Whyne (2018). Additionally, some literature suggests empirically determining the window length (Jalal et al., 2017).

After segmentation, a dimensional reduction procedure using principal components analysis is applied, retaining sensor data from both hands. Three window lengths, one-third, one-half, and two-thirds of the average length of training samples, are employed with a 75% overlap for window segmentation and optimal size determination. Zero padding, a standard method during insufficient sample length, is used, following recommendations (Hashemi, 2019; Um et al., 2017).

## Feature selection and machine learning model training

In the feature extraction phase, the study selects quaternions, three-axis accelerations, and three-axis angular velocities from the IMUs during the operation process to derive 11 commonly used time-domain features (see Table 1 for details).

Four machine learning algorithms – SVM, RF, NB, and LSTM – are applied in this research. NB is frequently used as a baseline for comparison with other algorithms in practice. The training and prediction time of NB is rapid because it is a parametric classifier. Many related studies, such as those by Sefen et al. (2016) and Liu et al. (2016), have used this classifier for HAR.

SVM is another popular parametric classifier. Its advantages include insusceptibility to outliers, as classification is determined by
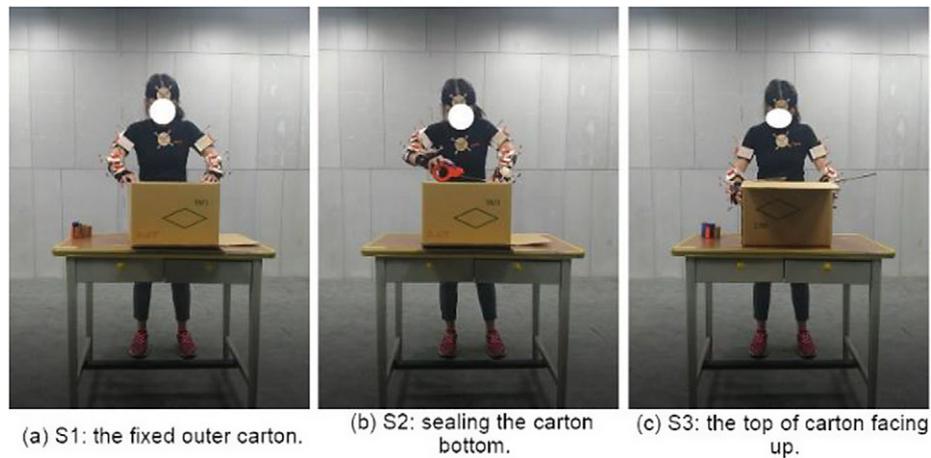
Figure 3. Operational elements decomposition of the assembly of outer carton (T2).
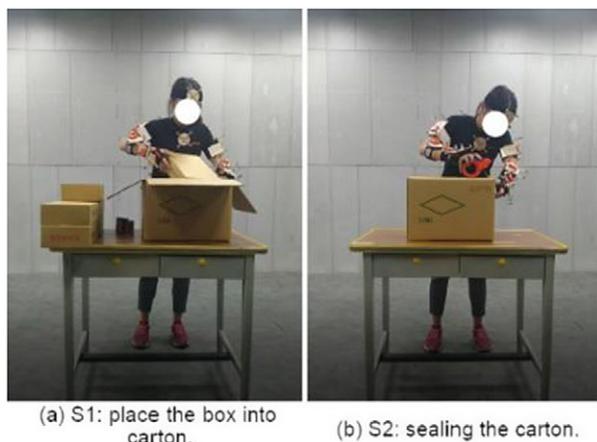
Figure 4. Operational elements decomposition of the final packing of the carton for shipment (T3).

Table 1. Time-domain features applied in this study

Description
The average of each variable
The median value of each variable
The sum of square of each variable
The standard deviation of each variable
The variance of each variable
The minimum value of each variable
The maximum of each variable
The skew of each variable
The Kurt of each variable
The mean square error of each variable
The number of mean crossings for each variable

RF is a classification model composed of many decision trees, based on ensemble learning. The final output class of the random forest is determined by the majority vote of the output classes of the many decision trees, which helps avoid the overfitting problem. The random forest algorithm often achieves the highest accuracy in HAR literature, as seen in studies by Baldominos et al. (2019), Xu et al. (2017), and Chetty and White (2016).
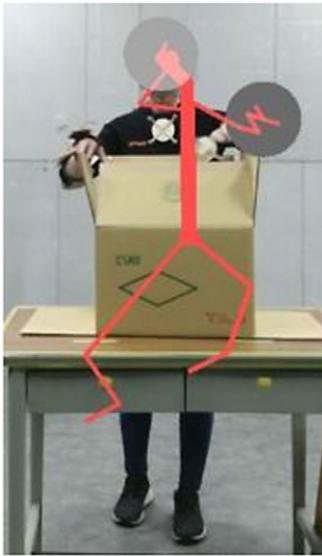
In recent years, deep neural networks with more than four layers have sparked a new wave of AI research and applications in deep learning. Among the various deep learning models, CNNs imitate the neural circuits of the visual processing area of the human brain, enabling them to recognize and classify objects in static images with excellent performance. RNNs, a type of neural network that processes sequential data such as sound, language, and video, can use previous outputs as inputs for the next time step, making them a neural network with short-term memory. LSTM improves upon RNN by using a conveyor belt-like memory line and multiple gates to filter and process data requiring long-term memory, making it similar to the hippocampus of the human brain, responsible for processing both short-term and long-term memory. The research topic of this study involves the decomposition and recognition of the movements of industrial operation personnel, which falls under the analysis of sequential data. In the literature, Pienaar and Malekian (2019) used LSTM for HAR and achieved excellent results. Therefore, this study chose LSTM as the representative method in deep learning.

Results

Due to cardboard boxes obstructing participants' bodies, the Microsoft Kinect V2 produced distorted postures (Fig. 5), leading to the exclusion of its kinematic data from this study. Therefore, only RGB images were retained for the analysis of misclassification by the machine learning models.

Data preprocessing

After normalizing the XSENS-collected variables, samples were segmented using time windows (TW) with lengths 1/3 (L1), 1/2 (L2), and 2/3 (L3) of the average training sample length, incorporating a 75% overlap. Table 2 displays the determined optimal time window size.

the support vector, and support for both linear and nonlinear separation through different kernel tricks. SVM is also commonly applied in HAR research, such as in studies by Burns et al. (2018) and Subasi et al. (2018), and has achieved good accuracy.

Figure 5. The Microsoft Kinect Skeleton diagram reveals significant distortions in body posture compared to the actual body position.

Table 2. The determined optimal time window size for tasks in this study

Tasks	TW L1	TW L2	TW L3
T1	143	215	287
T2	221	331	441
T3	530	795	1060

Table 3. The best model for the inner box assembly (T1)

Model	Normalization	WL	Accu.	F1	TT (s)
SVM	MA	287 (L3)	0.961	0.910	1.375
RF	MA	287 (L3)	0.933	0.873	2.078
NB	MM	287 (L3)	0.867	0.772	1.312
LSTM	MM	287 (L3)	0.683	0.494	305.847

Models training and implementation

The study divided the data into two parts: testing data reserved for participants P1–P3 to verify model performance and training data using the remaining participants for learning. During the experiment, the researchers observed that three participants (P1–P3) exhibited movements inconsistent with the standard actions we set, so their data were excluded from the training set to ensure the accuracy of the trained models. However, we utilized the imperfect data when testing our models to determine whether the models could recognize incorrect and inefficient operational elements performed by workers. The validation results showed that nonmatching elements were accurately identified. This provided strong evidence that our research objective of developing machine learning models to aid industrial engineers in improving workforce productivity on the shop floor was met.

Classification of operational elements in inner box assembly (T1)

In the assembly of inner boxes, four operational elements were identified. Posttraining, the optimal classification model for each algorithm was determined (Table 3), with all four algorithms achieving peak accuracy at a time window length (WL) of 287 frames (L3). SVM and RF exhibited improved accuracy (Accu.) with MA normalization, while NB and LSTM demonstrated enhanced accuracy with MM normalization. Notably, SVM attained the highest accuracy at 96.1% and an F1 score at 0.910, with a model training time (TT) only slightly exceeding 0.05 seconds compared to NB.

Figure 6 shows the confusion matrix of the SVM model classification results. Participants 1 and 3 accurately classified the operational elements. However, for Participant 2, the confusion matrix indicates that the actual left side of the assembled box (S2) was predicted as the operational element of closing the cover (S4).

Figures 7 and 8 display images of operational elements S2 and S4 performed by Participant 2. Participant 2 exhibited confusion and a lack of proficiency in these two actions. These figures illustrate that assembling the inner box relies heavily on the use both hands, requiring more delicacy than the other tasks. Therefore, when training the classification model, the weight of hand motion data in the assembly of the inner box task can be increased prior to training the model.

### Classification of operational elements in outer carton assembly (T2)

Table 4 illustrates the best classification model for assembling outer cartons for each algorithm. Except for LSTM, all other algorithms achieved identical accuracy and F1 scores with both normalization methods. When different algorithms achieved the same accuracy and F1 scores with varying normalization methods, the model with the shorter training time was selected as the best model. In this study, the NB algorithm had the fastest training time and achieved the highest accuracy of 97.9% with a window size of 441 frames (L3).

Figure 9 presents the confusion matrix of the classification results. Participant 2 experienced confusion between fixing the outer carton (S1) and turning the bottom of the carton up (S3). However, Participants 2 and 3 were accurately classified into their respective operational elements.

Figures 10 and 11 showcase the video frames of the two subsets of operational elements performed by Participant 1 while executing fixing the inner box (S1) and turning the bottom of the carton up (S3). For S3, participants were only required to rotate the carton from the bottom to the top so that the opening of the carton faced upward, finally of Participant 1 during the carton facing upward (S3), it can be observed that Participant 1 performs a movement of tidying up the opening of the carton. It is speculated that this additional movement contributed to the misclassification of the operational element.

### Classification of operational elements in packing task (T3)

Table 5 illustrates the top-performing model among various classification algorithms for the packing operation. Apart from LSTM, the other algorithms exhibited identical accuracy and F1 values with both normalization methods (see Table 5). Likewise, the model with the shorter training time was chosen as the optimal one. Throughout this study, all four algorithms achieved 100% accuracy and an F1 value of 1. Notably, both SVM and RF attained 100% accuracy and an F1 value of 1 across three different window lengths of training, with the results from the shortest window length being considered representative. While NB reached the highest accuracy solely at the window length of 795 frames (L2), it incurred additional time. Although LSTM did not achieve the highest accuracy and F1 value at window lengths 795 frames (L2) and 1060 frames
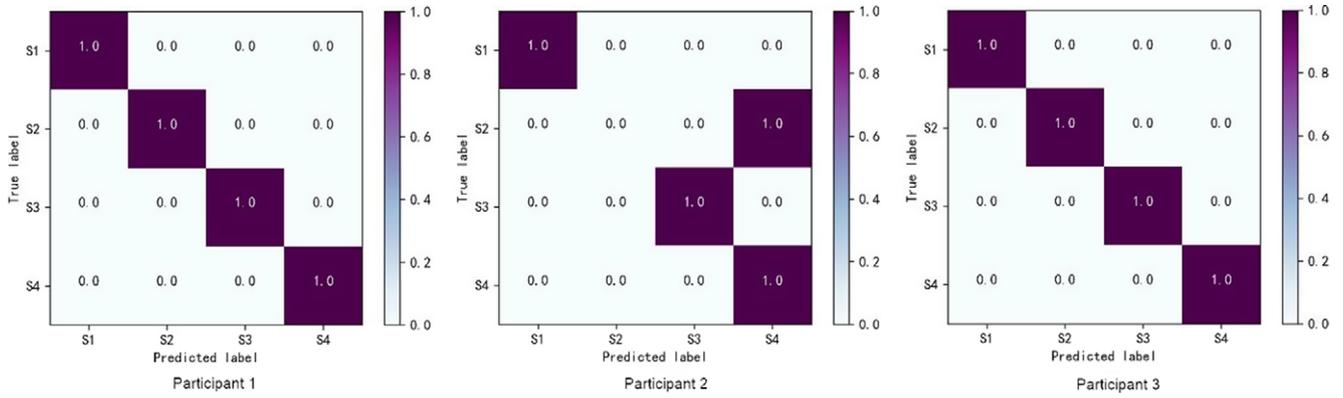
Figure 6. The confusion matrix of the best SVM model with MA normalization and a window width of 287 frames for T1 in the participant-independent validation phase.
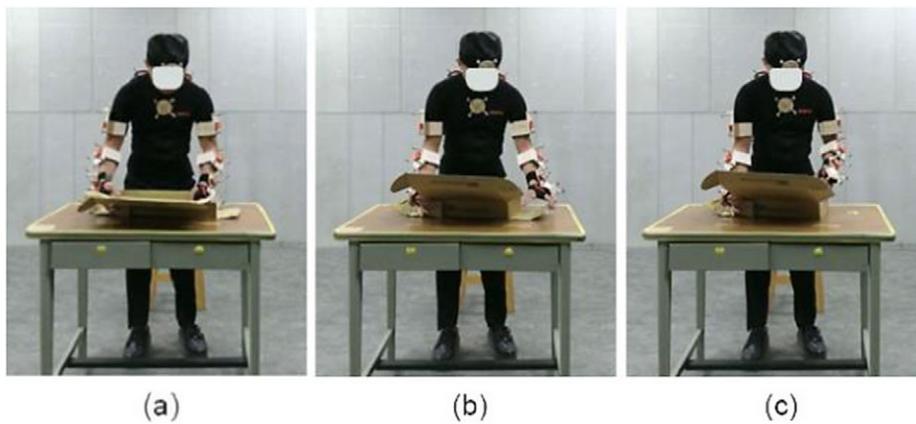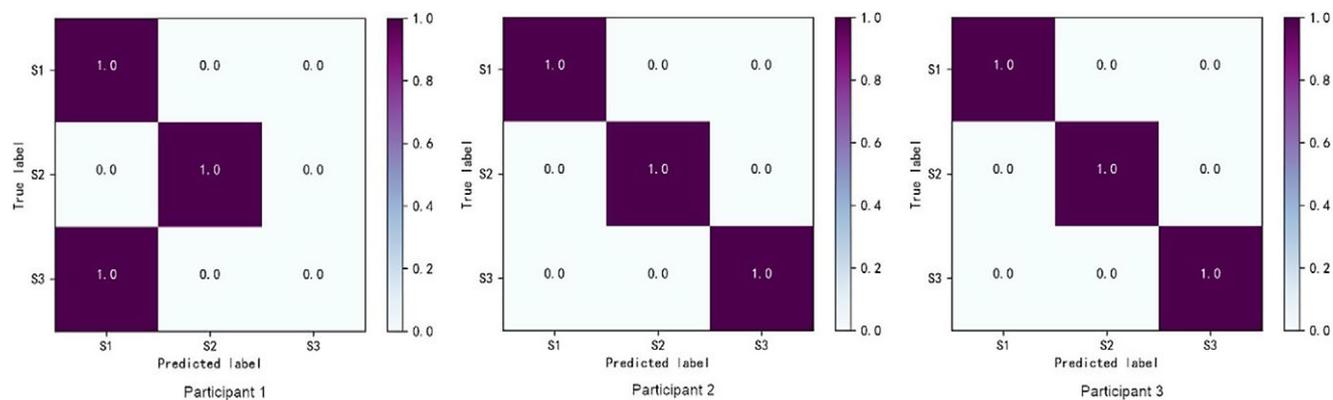

Figure 7. The video clip of Participant 2 performing the S2 element of the assembly of the inner box (T1) from (a) to (c) in order.
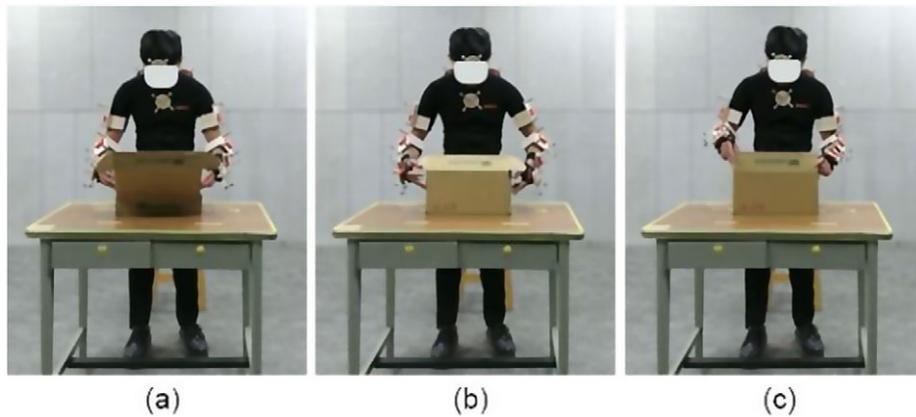
Figure 8. The video clip of Participant 2 performing the S4 element of the assembly of the inner box (T1) from (a) to (c) in order.

Table 4. The best model for the outer carton assembly (T2)

Model	Normalization	WL	Accu.	F1	TT (s)
SVM	MA	441 (L3)	0.979	0.880	1.656
RF	MM	441 (L3)	0.979	0.880	1.887
NB	MM	441 (L3)	0.979	0.880	1.625
LSTM	MM	441 (L3)	0.824	0.800	529.890

Figure 9. The confusion matrix of the best NB model with MM normalization and a window width of 441 frames for T2 in the participant-independent validation phase.
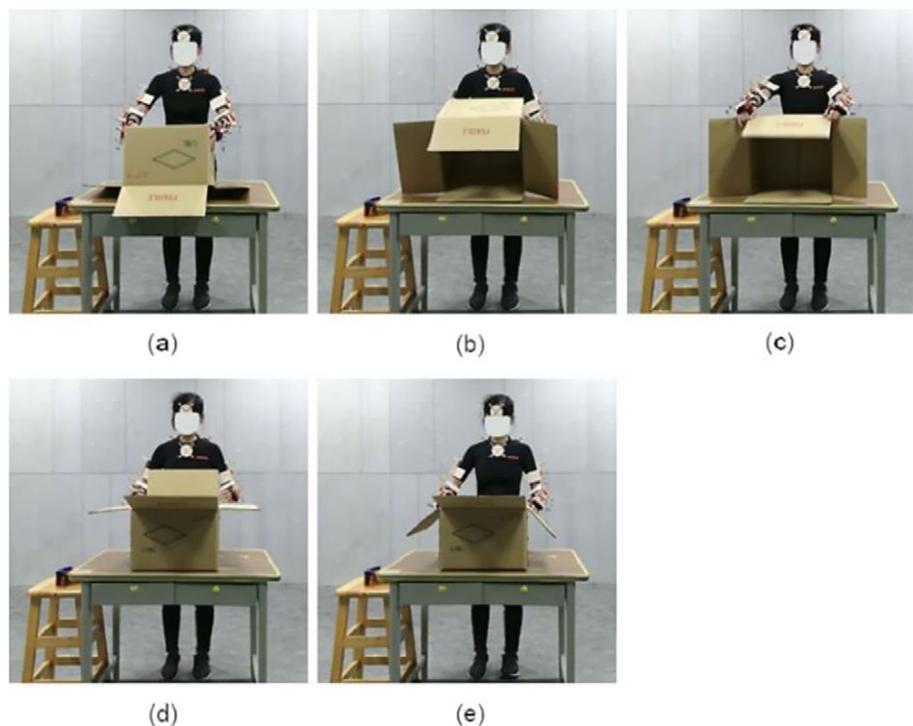
Figure 10. The video clip of Participant 1 performing the S1 element of the assembly of the outer carton (T2).

(L3), it matched the 100% accuracy and F1 value of 1 achieved by other algorithms at window length 530 frames (L1).

Figure 12 demonstrates the SVM model's accurate classification of the two operational elements in the packing operation – placing the inner box into the outer carton (S1) and sealing the carton (S2) – as performed by all three participants.

Discussion

Although this study also attempted to use vision-based devices to collect kinematic data, the highly distorted skeleton diagram in Figure 5 raised concerns about the reliability of the data. We found that when an object is positioned between the vision-based device and the operator, accurate data cannot be obtained. This result is consistent with findings in the literature (Vrigkas et al., 2015). However, RGB images are helpful in analyzing misclassifications

(see Figures 7, 8, 10, and 11). Therefore, sensor-based and vision-based devices should be used together whenever possible.

Table 6 compares the best performance between the literature and the current study. Several factors can affect the performance of prediction models. The first factor is the type of study. In theory, the prediction accuracy of participant-dependent studies tends to be higher than that of participant-independent studies because in participant-dependent studies, the training and testing data come from the same participants. In other words, the trained model has already "seen" the data. Therefore, in Table 6, the accuracy of participant- dependent studies is nearly perfect, except for the study by Beli et al. (2018). If the data come from a small number of participants or even just one (Lucci et al., 2022; Wang and Santoso, 2022), the results are also outstanding. After all, the more participants there are, the greater the variability, and thus, accuracy decreases.
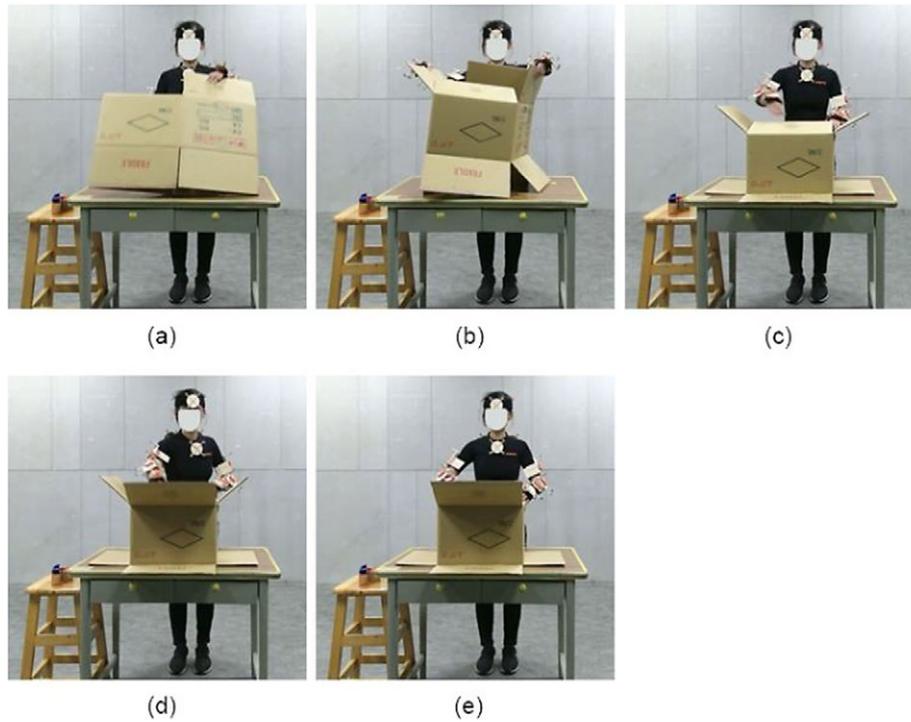
Figure 11. The video clip of Participant 1 performing the S3 element of the assembly of the outer carton (T2).

Table 5. The best model for the packing task (T3)

Model	Normalization	WL	Accu.	F1	TT (s)
SVM	MM	530 (L1)	1.000	1.000	1.883
RF	MM	530 (L1)	1.000	1.000	2.042
NB	MM	795 (L2)	1.000	1.000	2.125
LSTM	MM	530 (L1)	1.000	1.000	1330.204
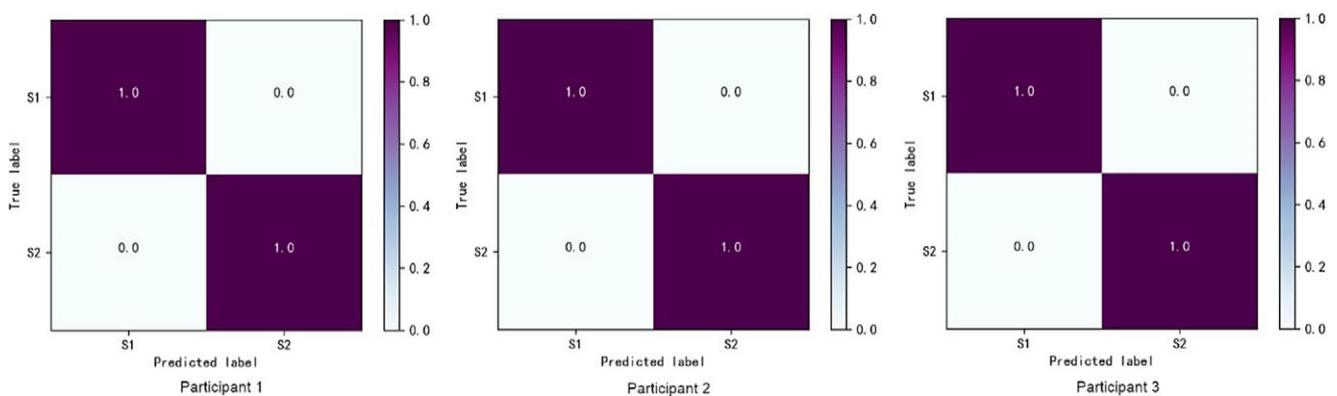
Figure 12. The confusion matrix of the best SVM model with MM normalization and a window width of 530 frames for T3 in the participant-independent validation phase. All predictions are correct.

The second factor is the degree of distinction between the classification targets. For example, Ishibashi and Fujii (2021) studied the lifting of a box from the floor to a workbench, while Li et al. (2022) focused on martial arts movements. These actions are larger and more distinct, unlike the fine hand movements targeted in this study, which are more difficult to distinguish.

The third factor is the algorithm. Lucci et al. (2022) utilized the logic of the positions of the workpiece and the operator's hands, along with the sequence of the machine's operations, rather than a machine learning algorithm. Although the programming is more complex, if the algorithm is well-designed, the accuracy can also be nearly perfect.

Table 6. The comparison of the best models in the literature and the current study

Study	Target	Sector	Device	Number of elements	Accuracy (%)	Algorithm	Type	Number of participants
Beli et al., 2018	EE	ADL	IMU	9	84.5	HMM-based	PD	10
Ishibashi and Fujii, 2021	EE	Ind.	IMU	3	99.8	HMM-based	PI	29 + 1
Li et al., 2022	EE	SP	IMU	8	99.6	DTW-based	PD	54
Tuli et al., 2022	EE	Ind.	RGB + IMU	9	88.0	CNN + LSTM	PI	Unknown[a] + 1
Jonek et al., 2023	EE	Ind.	RGB + IMU	5	76.0	CNN + LSTM	PI	Unknown[a] + 1
Lucci et al., 2022	ETS	Ind.	RGB	7	99.3	Logic	PD	1
Wang and Santoso, 2022	ETS	Ind.	RGB	5	95.1	YOLO v3	PD	1
This study	EE	Ind.	IMU	4 (T1)	96.1	SVM	PI	13 + 3
				3 (T2)	97.9	SVM, RF, NB		
				2 (T3)	100	SVM, RF, NB		

[a]The number of participants for the model development was not reported.

Note: EE = entire element; ETS = elements in timestamp; ADL = activity of daily living; Ind. = industrial operation; SP = sport; HMM = hidden Markov model; DTW = dynamic time warping; Logic = applying pure logic but not any machine learning model; YOLO = you only look once; PD = participant-dependent; PI = participant-independent.

In the three tasks of this study, the hand movements were smallest in the inner box assembly, larger in the outer carton assembly, and largest in the packing task. Consequently, the degree of distinction between the movements increased from low to high. This result also explains the differences in accuracy shown in Tables 3, 4, and 5.

Another interesting finding of this study is that the performance of simple, traditional machine learning models was better than that of the deep learning LSTM model (Tables 3, 4, and 5). The training time for LSTM exceeded 1,000 seconds, whereas the training times for SVM, RF, and NB were all under 3 seconds. Moreover, SVM achieved the highest accuracy, with a minimum value of 0.961 occurring in the task with the lowest distinction, the inner box assembly (Table 3). Although the accuracy of RF and NB was slightly lower in the inner box assembly, they performed as well as SVM in tasks with higher distinction, such as the outer carton assembly (0.979 in Table 4) and the packing task (1.000 in Table 5). Our findings for SVM, RF, and NB align with the results in the literature (Baldominos et al., 2019; Burns et al., 2018; Chetty and White, 2016; Liu et al., 2016; Sefen et al., 2016; Subasi et al., 2018; Xu et al., 2017).

Although the accuracy of LSTM in this study was not as high as that of SVM, RF, and NB, its accuracy across the three tasks was 0.683, 0.800, and 1.000, respectively. Compared with the two studies in Table 6 that used CNN for feature extraction and LSTM as the classifier – 760 in Jonek et al. (2023) and 0.880 in Tuli et al. (2022) – these results are quite comparable. However, the reason for LSTM's relatively lower accuracy remains unknown and requires further investigation in future studies.

The results of this study (Tables 3, 4, and 5) demonstrate that operational motion analysis can be performed without expert knowledge and significant manual effort. This approach provides insights into the process, which can be used to identify novices' awkward operations (Figures 7, 8, 10, and 11). Therefore, HAR has the potential for optimization to increase productivity, supporting the argument of Jonek et al. (2023), and can help junior workers follow SOPs, consistent with the claim of Wang and Santoso (2022).

Recently, Industry 5.0, inheriting all the advanced technologies from Industry 4.0 (Iqbal et al., 2022), emphasizes the vision of building sustainable, human-centric, and resilient industries. The critical focus of Industry 5.0 is reintroducing human workers to the shop floors, which has attracted significant interest from researchers (Alves et al., 2023; Cimino et al., 2023). The primary objective is to enhance human creativity and intelligence in smart manufacturing systems (Iqbal et al., 2022), while ensuring the well-being of workers (Moller et al., 2022). Therefore, the application of time series-based HAR in human–robot collaboration is crucial for Industry 5.0, and the emergence of related research (Lucci et al., 2022; Wang and Santoso, 2022) aligns with this trend. However, future studies in this domain should consider using a more diverse set of participants to develop and validate models.

Conclusions

The primary objective of this study was to develop machine learning models for HAR tailored to manufacturing tasks, diverging from traditional HAR in ADL. Focusing exclusively on recognizing detailed operational elements provides a crucial foundation for enhancing work efficiency. To achieve this goal, the machine learning model must identify the specific operating elements of tasks performed by workers rather than the entire operation.

The training dataset from three benchmark workers consists of thousands of data points with specific motion trajectory patterns, ensuring the trained model's robustness to recognize movements performed by different participants. With machine learning models capable of recognizing sequential motions, industrial engineers in Industry 5.0 can adopt intelligent methods to replace traditional, manual motion and time study approaches, enabling more rapid and objective improvements in production efficiency.

This study replicated shipping logistics operations in a semiconductor design house to build, train, and validate machine learning models. Results indicate that

1. SVM, RF, and NB machine learning models developed in this study accurately classify various operational elements in three tasks using time window segmentation, overlapping, and normalization techniques.
2. In cases of misclassifications, RGB clips can be used to scrutinize the effectiveness and correctness of employee movements. This enables supervisors to devise targeted training plans and assists industrial engineers in conducting motion and time studies with modern tools. The method proposed in

this study can improve the efficiency of industrial engineers and on-site supervisors, reduce workloads, and mitigate conflicts stemming from workers feeling surveilled and dissatisfaction from feeling disrespected.

Despite favorable outcomes, there are limitations:

1. Although the research meets the theoretical functional requirements of traditional industrial engineering in motion and time studies, a more detailed decomposition of operational elements may be necessary for realistic applications in today's highly complex manufacturing tasks. For example, the current research decomposed a single operation into only four simple assembly steps, but in actual industrial settings, these hand operations are often broken down into more than 10 subactions, with a need to differentiate between left- and right-hand movements.

2. Operation completion time is an objective indicator of work efficiency improvement. Comparing the completion time of each step to the standard operation time would be useful and should be investigated in future research.

3. The decomposition of motion elements still relies on human intervention. Future integration of unsupervised learning may cluster motions of standard workers into operational elements for each task, reducing the workload on industrial engineers.

References

Aggarwal JK and Ryoo MS (2011) Human activity analysis. *ACM Computing Surveys* **43**(3), 1–43. https://doi.org/10.1145/1922649.1922653

Aggarwal JK and Xia L (2014) Human activity recognition from 3D data: A review. *Pattern Recognition Letters* **48**, 70–80. https://doi.org/10.1016/j.patrec.2014.04.011

Al-Amin M, Qin R, Tao W, Doell D, Lingard R, Yin Z, and Leu MC (2020) Fusing and refining convolutional neural network models for assembly action recognition in smart manufacturing. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* **236**, 2046–2059.

Alves J, Lima TM and Gaspar PD (2023) Is Industry 5.0 a human-centred approach? *A systematic review.* Processes **11**(1), 193. https://doi.org/10.3390/pr11010193

Anguita D, Ghio A, Oneto L, Parra X and Reyes-Ortiz JL (2013) A public domain dataset for human activity recognition using smartphones. In *Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning.* Bruges, Belgium: ESANN, p.3.

Argyriou V, Petrou M and Barsky S (2010) Photometric stereo with an arbitrary number of illuminants. *Computer Vision and Image Understanding* **114**(8), 887–900. https://doi.org/10.1016/jxviu.2010.05.002

Aurand AM, Dufour JS and Marras WS (2017) Accuracy map of an optical motion capture system with 42 or 21 cameras in a large measurement volume. *Journal of Biomechanics* **58**, 237–240. https://doi.org/10.1016/jjbiomech.2017.05.006

Baldominos A, Cervantes A, Saez Y and Isasi P (2019) A comparison of machine learning and deep learning techniques for activity recognition using mobile devices. *Sensors (Basel)* **19**(3), 521. https://doi.org/10.3390/s19030521

Banos O, Garcia R, Holgado-Terriza JA, Damas M, Pomares H, Rojas I, Saez A and Villalonga C (2014) mHealthDroid: A novel framework for agile development of mobile health applications. In Pecchia L, Chen LL, Nugent C and Bravo J (eds), *Ambient Assisted Living and Daily Activities: 6th International Work-Conference (IWAAL 2014)*, December 2–5, Proceedings 6. Belfast, UK: Springerpp., 91–98.

Beil J, Ehrenberger I, Scherer C, Mandery C and Asfour T (2018) Human motion classification based on multi-modal sensor data for lower limb exoskeletons. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* Madrid, Spain: IEEE, 5431–5436.

Bozkurt F (2021) A comparative study on classifying human activities using classical machine and deep learning methods. *Arabian Journal for Science and Engineering* **47**(2), 1507–1521. https://doi.org/10.1007/s13369-021-06008-5

Bulling A, Blanke U and Schiele B. (2014) A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys* **46**(3), 1–33. https://doi.org/10.1145/2499621

Burns DM, Leung N, Hardisty M, Whyne CM, Henry P and McLachlin S (2018) Shoulder physiotherapy exercise recognition: Machine learning the inertial signals from a smartwatch. *Physiological Measurement* **39**(7), 075007.

Burns DM and Whyne CM (2018) Seglearn: A python package for learning sequences and time series. *The Journal of Machine Learning Research* **19**(1), 3238–3244.

Chavarriaga R, Sagha H, Calatroni A, Digumarti ST, Tröster G, Millán JdR and Roggen D (2013) The opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters* **34**, 2033–2042.

Chetty G and White M (2016) Body sensor networks for human activity recognition. Paper presented at the 3rd International Conference on Signal Processing and Integrated Networks (SPIN), 11–12 February, Amity School of Engineering and Technology, Noida, India.

Chien C-F, Hong T-y and Guo H-Z (2017). A conceptual framework for "Industry 3.5" to empower intelligent manufacturing and case studies. *Procedia Manufacturing* **11**, 2009–2017. https://doi.org/10.1016/j.promfg.2017.07.352

Cimino A, Elbasheer M, Longo F, Nicoletti L and Padovano A (2023) Empowering field operators in manufacturing: A prospective towards Industry 5.0. *Procedia Computer Science* **217**, 1948–1953.

Demir KA, Döven G and Sezen B (2019) Industry 5.0 and human-robot co-working. *Procedia Computer Science* **158**, 688–695.

Dua N, Singh SN, Challa SK, Semwal VB and Sai Kumar MLS (2022) A survey on human activity recognition using deep learning techniques and wearable sensor data. In Khare N, Tomar DS, Ahirwal MK, Semwal VB and Soni V (eds), *Machine Learning, Image Processing, Network Security and Data Sciences.* Cham: Springer, pp. 52–71. https://doi.org/10.1007/978-3-031-24352-3_5

Fortino G, Galzarano S, Gravina R and Li W (2015) A framework for collaborative computing and multi-sensor data fusion in body sensor networks. *Information Fusion* **22**, 50–70.

Freivalds A (2009*)* Niebel's Methods, Standards, and Work Design. Boston, MA: Mcgraw-Hill Higher Education.

Gan ZL, Musa SN and Yap HJ (2023) A review of the high-mix, low-volume manufacturing industry. *Applied Sciences* **13**(3), 1687. https://doi.org/10.3390/app13031687

Hashemi M (2019) Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation. *Journal of Big Data* **6**(1), 1–13.

Hassan MM, Uddin MZ, Mohamed A and Almogren A (2018) A robust human activity recognition system using smartphone sensors and deep learning. *Future Generation Computer Systems* **81**, 307–313. https://doi.org/10.1016/jiuture.2017.11.029

He H, Tan Y and Zhang W (2018) A wavelet tensor fuzzy clustering scheme for multi-sensor human activity recognition. *Engineering Applications of Artificial Intelligence* **70**, 109–122. https://doi.org/10.1016/j.engappai.2018.01.004

Ignatov A (2018) Real-time human activity recognition from accelerometer data using convolutional neural networks. *Applied Soft Computing* **62**, 915–922. https://doi.org/10.1016/j.asoc.2017.09.027

Iqbal M, Lee CKM and Ren JZ (2022) Industry 5.0: From manufacturing industry to sustainable society. In *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM).* Kuala Lumpur, Malaysia: IEEE, 1416–1421.

Ishibashi N and Fujii F (2021) Hidden Markov model-based human action and load classification with three-dimensional accelerometer measurements. *IEEE Sensors Journal* **21**(5), 6610–6622. https://doi.org/10.1109/jsen.2020.3042201

Jalal A, **Kim Y-H**, **Kim Y-J**, **Kamal S and Kim D** (2017) Robust human activity recognition from depth video using spatiotemporal multi-fused features. *Pattern Recognition* **61**, 295–308. https://doi.org/10.1016/j.patcog.2016.08.003

Janidarmian M, **Roshan Fekr A**, **Radecka K and Zilic Z** (2017) A comprehensive analysis on wearable acceleration sensors in human activity recognition. *Sensors (Basel)* **17**(3), 529. https://doi.org/10.3390/s17030529

Johri P, **Singh JN**, **Sharma A and Rastogi D** (2021) Sustainability of coexistence of humans and machines: An evolution of Industry 5.0 from Industry 4.0. In *10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, pp. 410–414

Jonek M, **Tuli TB and Manns M** (2023) A motion capture-based approach to human work analysis for industrial assembly workstations. In Galizia FG and Bortolini M (eds), *Proceedings of the Changeable, Agile, Reconfigurable and Virtual Production Conference and the World Mass Customization & Personalization Conference*. Cham: Springer.

Khan IU, **Afzal S and Lee JW** (2022) Human activity recognition via hybrid deep learning based model. *Sensors (Basel)* **22**(1), 323. https://doi.org/10.3390/s22010323

Konar A and Saha S (2018). Gesture recognition. In *Principles, Techniques and Applications*. Cham: Springer.

Kong XT, **Yang X**, **Huang GQ and Luo H** (2018) The impact of industrial wearable system on Industry 4.0. In *IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*. Zhuhai, China: IEEE, pp. 1–6.

Kulsoom F, **Narejo S**, **Mehmood Z**, **Chaudhry HN**, **Butt A and Bashir AK** (2022) A review of machine learning-based human activity recognition for diverse applications. *Neural Computing and Applications* **34**(21), 18289–18324. https://doi.org/10.1007/s00521-022-07665-9

Li H, **Khoo S and Yap HJ** (2022) Implementation of sequence-based classification methods for motion assessment and recognition in a traditional Chinese sport (Baduanjin). *International Journal of Environmental Research and Public Health* **19**(3), 1744. https://doi.org/10.3390/ijerph19031744

Liu H and Wang L (2018) Gesture recognition for human-robot collaboration: A review. *International Journal of Industrial Ergonomics* **68**, 355–367.

Liu Y, **Nie L**, **Liu L and Rosenblum DS** (2016) From action to activity: Sensor-based activity recognition. *Neurocomputing* **181**, 108–115.

Longo F, **Padovano A and Umbrello S** (2020) Value-oriented and ethical technology engineering in Industry 5.0: A human-centric perspective for the design of the factory of the future. *Applied Sciences* **10**(12), 4182.

Lucci N, **Monguzzi A**, **Zanchettin AM and Rocco P** (2022) Workflow modelling for human–robot collaborative assembly operations. *Robotics and Computer-Integrated Manufacturing* **78**, 102384.

Maddikunta PKR, **Pham Q-V**, **Prabadevi B**, **Deepa N**, **Dev K**, **Gadekallu TR**, **Ruby R and Liyanage M** (2022) Industry 5.0: A survey on enabling technologies and potential applications. *Journal of Industrial Information Integration* **26**, 100257. https://doi.org/10.1016/j.jii.2021.100257

Martímez-Villaseñor L and Ponce H (2019) A concise review on sensor signal acquisition and transformation applied to human activity recognition and human–robot interaction. *International Journal of Distributed Sensor Networks* **15**(6), 1550147719853987. https://doi.org/10.1177/1550147719853987

Micucci D, **Mobilio M and Napoletano P** (2017) UniMiB SHAR: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences* **7**(10), 1101.

Moller DPF, **Vakilzadian H and Haas RE** (2022) From Industry 4.0 towards Industry 5.0. In *IEEE International Conference on Electro Information Technology (eIT)*, pp. 61–68.

Montini E, **Cutrona V**, **Gladysz B**, **Dell'Oca S**, **Landolfi G and Bettoni A** (2022) A methodology to select wearable devices for Industry 5.0 applications. In *IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 1–4.

Nahavandi S (2019) Industry 5.0—A human-centric solution. *Sustainability* **11**(16), 4371. https://doi.org/10.3390/su11164371

Papetti A, **Gregori F**, **Pandolfi M**, **Peruzzini M and Germani M** (2020) A method to improve workers' well-being toward human-centered connected factories. *Journal of Computational Design and Engineering* **7**, 630–643.

Pienaar SW and Malekian R (2019) Human activity recognition using LSTM-RNN deep neural network architecture. Paper presented at the 2019 IEEE 2nd Wireless Africa Conference (WAC), 18–20 August, Pretoria, South Africa.

Poppe R (2010) A survey on vision-based human action recognition. *Image and Vision Computing* **28**(6), 976–990. https://doi.org/10.1016/jjmavis.2009.11.014

Qi W, **Wang N**, **Su H and Aliverti A** (2022) DCNN based human activity recognition framework with depth vision guiding. *Neurocomputing* **486**, 261–271.

Ramanathan M, **Yau W-Y and Teoh EK** (2014) Human action recognition with video data: Research and evaluation challenges. *IEEE Transactions on Human-Machine Systems* **44**, 650–663.

Reiss A and Stricker D (2012) Introducing a new benchmarked dataset for activity monitoring. In *16th International Symposium on Wearable Computers*, pp. 108–109.

Reyes-Ortiz J-L, **Oneto L**, **Samà A**, **Parra X and Anguita D** (2016) Transition-aware human activity recognition using smartphones. *Neurocomputing* **171**, 754–767. https://doi.org/10.1016/j.neucom.2015.07.085

Romero D, **Bernus P**, **Noran O**, **Stahre J and Fast-Berglund Å** (2016) The Operator 4.0: Human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems. In *Advances in Production Management Systems. Initiatives for a Sustainable World*. Cham: Springer, pp. 677–686.

Romero D, **Stahre J and Taisch M** (2020) The Operator 4.0: Towards socially sustainable factories of the future. *Computers & Industrial Engineering* **139**, 106128. https://doi.org/10.1016/j.cie.2019.106128

San-Segundo R, **Echeverry-Correa JD**, **Salamea C and Pardo JM** (2016) Human activity monitoring based on hidden Markov models using a smartphone. *IEEE Instrumentation & Measurement Magazine* **19**(6), 27–31.

Sefen B, **Baumbach S**, **Dengel A and Abdennadher S** (2016) Human activity recognition. Paper presented at the Proceedings of the 8th International Conference on Agents and Artificial Intelligence, 24–26 February, Science and Technology Publications, Ltd., Rome, Italy.

Shoaib M, **Bosch S**, **Incel OD**, **Scholten H and Havinga PJ** (2016) Complex human activity recognition using smartphone and wrist-worn motion sensors. *Sensors (Basel)* **16**(4), 426. https://doi.org/10.3390/s16040426

Shotton J, **Sharp T**, **Kipman A**, **Fitzgibbon A**, **Finocchio M**, **Blake A**, **Cook M and Moore R** (2013) Real-time human pose recognition in parts from single depth images. *Communications of the ACM* **56**(1), 116–124.

Stisen A, **Blunck H**, **Bhattacharya S**, **Prentow TS**, **Kjærgaard MB**, **Dey A**, **Sonne T and Jensen MM** (2015) Smart devices are different. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pp. 127–140.

Subasi A, **Radhwan M**, **Kurdi R and Khateeb K** (2018) IoT based mobile healthcare system for human activity recognition. Paper presented at the 15th Learning and Technology Conference (L&T), 25–26 February, Effat University, Jeddah, Saudi Arabia.

Sztyler T and Stuckenschmidt H (2016) On-body localization of wearable devices: An investigation of position-aware activity recognition. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–9.

Tuli TB, **Patel VM and Manns M** (2022) Industrial human activity prediction and detection using sequential memory networks. *In* 3rd Conference on Production Systems and Logistics, pp. 62–72.

Um TT, **Babakeshizadeh V and Kulić D** (2017) Exercise motion classification from large- scale wearable sensor data using convolutional neural networks. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2385–2390.

Vrigkas M, **Nikou C and Kakadiaris IA** (2015) A review of human activity recognition methods. *Frontiers in Robotics and AI* **2**, 28. https://doi.org/10.3389/frobt.2015.00028

Wang J, **Chen Y**, **Hao S**, **Peng X and Hu L** (2019) Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters* **119**, 3–11. https://doi.org/10.1016/j.patrec.2018.02.010

Wang J., **Zhang X**, **Gao Q**, **Ma X**, **Feng X and Wang H** (2017) Device-free simultaneous wireless localization and activity recognition with wavelet feature. *IEEE Transactions on Vehicular Technology* **66**(2), 1659–1669. https://doi.org/10.1109/tvt.2016.2555986

Wang K-J and Santoso D (2022) A smart operator advice model by deep learning for motion recognition in human–robot coexisting assembly line. *The International Journal of Advanced Manufacturing Technology* **119**, 865–884.

Weinland D, **Ronfard R and Boyer E** (2011) A survey of vision-based methods for action representation, segmentation and recognition. *Computer Vision and Image Understanding* **115**, 224–241.

Xu L, **Yang W**, **Cao Y and Li Q** (2017) Human activity recognition based on random forests. In *13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 548–553.

Yadav SK, **Tiwari K**, **Pandey HM and Akbar SA** (2021) Skeleton-based human activity recognition using ConvLSTM and guided feature learning. *Soft Computing* **26**, 877–890.

Zappi P, **Lombriser C**, **Stiefmeier T**, **Farella E**, **Roggen D**, **Benini L and Tröster G** (2008) Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection. In *Wireless Sensor Networks: 5th European Conference (EWSN 2008)*, January 30–February 1. Bologna, Italy: Springer, pp. 17–33.

Zhu C and Sheng W (2011) Motion- and location-based online human daily activity recognition. *Pervasive and Mobile Computing* **7**(2), 256–269. https://doi.org/10.1016/j.pmcj.2010.11.004