Published: 10 November 2025

Citation in BibTeX format

CIKM '25: The 34th ACM International Conference on Information and Knowledge Management
November 10 - 14, 2025
Seoul, Republic of Korea

Conference Sponsors:
SIGWEB
SIGIR

SHORT-PAPER

Internet of Things Dataset for Human Operator Activity Recognition in Industrial Environment

ABDUR RAHIM MOHAMMAD FORKAN, Swinburne University of Technology, Hawthorn, VIC, Australia

PREM PRAKASH JAYARAMAN, Swinburne University of Technology, Hawthorn, VIC, Australia

CLARENCE ANTONMERYL, Swinburne University of Technology, Hawthorn, VIC, Australia

FEDERICO MONTORI, University of Bologna, Bologna, BO, Italy

ABHIK BANERJEE, Swinburne University of Technology, Hawthorn, VIC, Australia

KANEEZ FIZZA, Swinburne University of Technology, Hawthorn, VIC, Australia

View all

Open Access Support provided by:

Swinburne University of Technology

University of Bologna

CIKM '25: Proceedings of the 34th ACM International Conference on Information and Knowledge Management (November 2025)
https://doi.org/10.1145/3746252.3761649
ISBN: 9798400720406

Internet of Things Dataset for Human Operator Activity Recognition in Industrial Environments

Abdur Forkan
Swinburne University
Melbourne, Australia
fforkan@swin.edu.au

Prem Prakash Jayaraman
Swinburne University
Melbourne, Australia
pjayaraman@swin.edu.au

Clarence Antonmeryl
Swinburne University
Melbourne, Australia
cantonmeryl@swin.edu.au

Federico Montori
University of Bologna
Bologna, Italy
federico.montori2@unibo.it

Abhik Banerjee
Swinburne Universit
Melbourne, Australia
abanerjee@swin.edu.au

Kaneez Fizza
Swinburne University
Melbourne, Australia
kfizza@swin.edu.au

Dimitrios Georgakopoulos
Swinburne University
Melbourne, Australia
dgeorgakopoulos@swin.edu.au

Abstract

In industrial environments, most production-related activities performed by human operators are often complex. Accurate detections of these activities are pivotal as it can greatly help to assess productivity that can lead to improvement in worker training, as well as in other scenarios ensure a safe work environment and reducing injuries. Existing datasets on wearable Internet of Things (IoT) for human activity recognition primarily focuses on general activities, such as walking, running, etc., and therefore, related machine learning models and datasets are not suitable for application to industrial environments. In this paper, we present a novel dataset for classifying human operator activities in a meat processing plant where production line operators use knives to cut, process and produce meat products. Our dataset contains human operator activity data captured using wearable IoT sensors collected from a meat processing production facility. Through extensive experiments using machine and deep learning, we demonstrate that our dataset is effective and useful for detecting different activities of a human operator working in an industrial environment. To the best of our knowledge, this is the only real-world IoT dataset that will be made publicly available to support further research into industrial activities recognition. Our dataset and related experiments are available at https://digitalinnovationlab.github.io/mppdataset.

CCS Concepts

• **Computing methodologies → Distributed computing methodologies**; **Classification and regression trees**.

Keywords

Activity recognition, Wearable Internet of Things, Machine Learning, Deep Learning, Human Operator

ACM Reference Format:
Abdur Forkan, Prem Prakash Jayaraman ,, Clarence Antonmeryl, Federico Montori, Abhik Banerjee, Kaneez Fizza, and Dimitrios Georgakopoulos . 2025. Internet of Things Dataset for Human Operator Activity Recognition in Industrial Environments. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25), November 10–14, 2025, Seoul, Republic of Korea.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3746252.3761649

1 Introduction

Improving the productivity of industrial plants is one of the major challenges in manufacturing, as this directly affects plant profitability [8]. Digital manufacturing is being increasingly adopted in manufacturing sectors towards developing data-driven approaches to improve efficiency. While existing digital manufacturing approaches primarily focus on machine and product monitoring for applications such as predictive maintenance and defect detection, there is relatively less emphasis on assessing and monitoring human operators. Characterisation and assessment of operator actions can help improve productivity [3] and workplace environment, such as by detecting safe work conditions [16] (e.g., use of a blunt knife in the meat processing plants [8]) and enhance worker skills [11].

Monitoring operator actions necessitates recognising and measuring individual actions undertaken by human operators during production. Human Activity Recognition (HAR) has seen extensive research focus in the past, spanning multiple domains [17]. However, existing literature as well as publicly available HAR datasets primarily use wearables or smartphones to recognise regular human actions such as walking, standing, etc. [4], [2], [14]. Such datasets cannot be used to develop Artificial Intelligence (AI) approaches for HAR in manufacturing scenarios, which involve specialised skills and dexterity required for handling industrial tools in production environments. An example of this is the use of industrial knives in meat processing plants, wherein both productivity and operator safety depend upon actions such as cutting, sharpening, idle, etc.

There is a critical need for collecting datasets correlating IoT sensor data with human operator actions to enable the development of accurate AI models for HAR in manufacturing scenarios. While some publicly available datasets have been published in recent literature, these often make use of video data which have limited applicability to other environments, especially due to privacy concerns surrounding video use. Significantly, a key drawback of existing datasets is that the data collection relates to curated experiments conducted in controlled lab environments, thereby limiting their applicability to actual production environments.

In our previous work [8], we developed a HAR model for evaluating human operator performance by presenting a real-world

---

CIKM '25, Seoul, Republic of Korea

ACM ISBN 979-8-4007-2040-6/2025/11
https://doi.org/10.1145/3746252.3761649

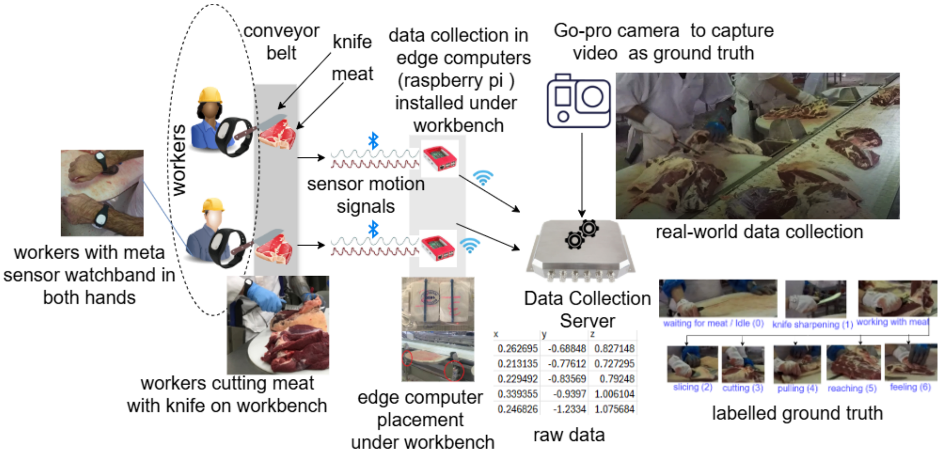

Figure 1: Real-world data collection setup from meat processing plant operators using wearable IoT sensor

use case in an Australian meat processing facility. In that work, we described our solution, the data collection process using wearable IoT and a preliminary machine learning (ML) model to detect human operator activities while using knives and alignment tools to cut, process, and produce meat products. In this work, we are publicly releasing our collected dataset, which includes both raw and pre-processed inertial sensor measurements, alongside extensive experimental results from various ML and deep learning (DL) models. The key contributions of this paper are highlighted as follows.

- We present a novel dataset for assessing human operator activities in industrial environments that involves knife handling. We released our data both in raw and pre-processed form (i.e., after conducting feature engineering) with annotated activities. The pre-processed data contains 2 versions: a simplified one containing 3 activity classes and a more detailed one containing 7 activity classes (Figure 1).
- We conducted extensive experiments with our data and presented the experimental outcomes of various ML and DL algorithms. We also released the developed ML and DL models for public use.
- We demonstrate that our dataset can lead to useful industrial applications for relevant industries, where human operators use knives for cutting (e.g., meat and food processing, poultry slaughterhouses, food and beverages, construction, textile manufacturing) and drive future research.

2 Related Work

In general, the data collection process for HAR in industrial scenarios can be classified into two categories: vision-based and IoT sensor-based. Vision-based solutions use different cameras to capture videos/images about the scenario, whereas IoT sensor-based solutions use wearable sensors, environmental sensors, or smartphones to collect data from a target (i.e., people/machine).

In the literature, vision-based methods are widely used in industrial setups for recognising operator activities. For example, InHard [6] released the data about activities related to Human-Robot collaboration that is collected from real-world settings. This industrial use case involves an assembly of various parts and components, and is carried out on different stages with the help of the robotic arm. Data is formed by videos in the format of RGD-S, containing over 2 million frames and collected from 16 distinct subjects. This dataset contains 13 different industrial action classes and over 4800 action samples. Authors in [7] created 10 industrial activity video datasets, each containing 3 types of actions. HRI130 [9] is another video-based dataset of Human-Robot interaction containing 30 categories of industrial-like actions and 2940 manually annotated clips. All these works mainly used DL (i.e., CNN or RNN-based methods) to classify industrial activities from video data. However, in general, vision-based systems have privacy concerns and data are sensitive to lighting, camera position, noise of vibrations, which are very common in industrial setups. In the case of IoT sensor-based solution, the authors in [15] employed a smartphone on operator's wrist and collected data of 5 industrial activities (e.g., Screwing with a screw driver, picking a part). The collected number of samples was 625 for 5 activities. Each sample contains 8 different signals, such as acceleration, brightness, and geometric field strength. The authors used DL to classify the activities. However, attaching a smartphone to the human body during production is not a viable solution to capture data in industrial settings, as it may affect the operator's performance. The use of environmental sensors to collect data is also not a suitable approach since environmental sensors are sensitive to the industrial environmental conditions and may not capture data with detailed granularity[13].

Wearable IoT (such as a watch-like sensor on the wrist) is the most convenient solution to perform HAR in industrial settings since it can be easily attached to the human body without much impacting the performance of operators. For instance, authors in [18] use Myo armband equipped with an inertial measurement unit (IMU) and sEMG sensors to collect activity data of assembly tasks. The data was collected from 8 subjects, however, they were in a controlled lab environment. Authors in [19] also used Myo armband to collect sEMG data of 9 key actions performed during onsite assembly and used DL to recognise those activities. However, these studies did not collect data in real-world settings, nor did they make their dataset publicly available.

3 The MPP Dataset

Our Meat Processing Plant (MPP) dataset was collected by deploying the solution described in [8] in one of Australia's largest meat processing plants. Human operators wore a watch-like IoT device (MetaWear watchband [8]) containing accelerometer and gyroscope sensor that capture the hand movements of operators while using the knife during normal production shifts. An overview of real-world data collection setup is shown in Figure 1. The architecture of our solution, the data collection and activity annotation process was described in [8].

3.1 About the Dataset

The data was collected from 4 human operators in 2 working days. The duration of the data collection was approximately 45 minutes for each operator. Overall, from 4 operators, a total of 11,629 seconds of clean data were extracted. The raw data comprises 16 CSV files (accelerometer and gyroscope data separately from 2 hands of 4 operators). There are 5 columns in each file. They are: time (the timestamp of the reading, starting from 0 in seconds), the x, y, and z acceleration/ gyroscope readings for the right/left hand, and class, which is the annotated activity observed in a reading. After combining data from both sensors of both hands in the same timestamp, the total number of datapoints in the raw data from 4 operators was 249280.

3.2 Raw data synchronisation

The raw data was collected from 2 Metawear sensors attached to the two hands of an operator with the same sampling rate. We aimed to use the raw data directly to train DL algorithms since this will give a higher number of data points which is suitable to train DL model. However, the timestamp of 2 sensors was not always synchronised. Moreover, there were some missing data points due to connectivity issues. To handle such an issue in the raw data, we merged the datapoints (accelerometer and gyroscope data from right and left hand) on timestamps to the closest signal within a certain tolerance value (0.01 or 0.02 seconds). For example, if there is one data point from the right hand at time 1.04 seconds and data from left hand at 1.05 seconds then they are matched under a tolerance of 0.01 and part of the same data instance. After that, each signal channel was standardised using StandardScaler. Total synchronised datapoints (from all operators) was 249280 for tolerance 0.01 and 84649 for tolerance 0.02. Afterwards, a sliding window technique is used with a fixed window size of 30 (based on sampling rate) and a stride of 15. This process increased the number of training examples for DL algorithms without altering the underlying data distribution. Each window captures a local temporal pattern, and the overlap ensures smooth transitions between segments.

3.3 Feature extraction

The features are derived from the accelerometer and gyroscope data on the x, y, and z axes for the left and right hands. For each measurement, the magnitude of accelerations and gyroscope were computed at the xyz, xy, yz, and xz planes. For these 7 values, the mean, standard deviation (std), area under the curve (AUC), maximum, and number of peaks over the records for one second were calculated, yielding 7 x 5 + 7 x 5 = 70 features. From the accelerometer data, the pitch and roll angles were also determined, and for all records over a second, the mean, std, AUC, and maximum were recorded, yielding another 4 + 4 = 8 features. This produces 78 features for each hand and 156 features for both hands.

3.4 Human operator activity labelling

We created two feature-engineered versions of the data namely 'class_3_dataset' (AD3) and 'class_7_dataset' (AD7). Both datasets have 157 columns (156 pre-processed features as described in the previous section and 1 activity label) that resulted 11,629 data instances. AD3 is labelled with 3 activities (idle, working and knife sharpening), whereas in AD7 working activity is expanded to 5 different activities, resulting in a total of 7 activities (idle, knife sharpening, slicing, cutting, pulling, reaching, and feeling) [8]. These activities are shown in Figure 1. The raw data are also annotated with 3 and 7 classes of activities. Most of the data was dominated by working (2) activities and within working activities, a majority of the data points were slicing (2), with little representation in the feeling (6) and pulling (4) activities. That is, certain classes are quite under-represented. With 3 classes, class 0 and 1 have a lower number of samples, and with 7 classes, class 4 and 6 are under-represented. Therefore, this became an imbalanced dataset.

3.5 Data Augmentation on pre-processed data

To overcome the class imbalance issue, the data is normalised and synthetic samples is created using SMOTE [5]. This creates synthetic samples for minority classes by selecting a sample, identifying its nearest neighbours, and creating new data points along the lines connecting the sample to its neighbours. This balances class distributions, reducing bias towards the majority class.

3.6 Dimensionality reduction on processed data

It is observed from correlation analysis that several features are highly correlated with each other, making them redundant. Moreover, features can be streamlined by removing those that are not highly correlated with the class. Therefore, we used Principal Component Analysis (PCA) [1] to transform the high-dimensional dataset into a lower-dimensional form while preserving as much variance as possible by reducing the number of features.

4 Experiments

In this section, the results of the performance evaluation for different ML and DL models trained on 3-class and 7-class activity data are presented to demonstrate the value of our dataset. The evaluation outcome is also valuable to measure the real-time efficiency of human operators while working in real industry environments.

4.1 Performance evaluation using ML models

We used 6 ML models that are trained on data with computed features (AD3 and AD7). The models are picked based on different varieties of their underlying structure (Table 1).

To evaluate the performance of our pre-processed dataset containing features with different ML models, we conducted separate experiments for AD3 and AD7 datasets. The results are presented in Table 1 and 2. We present the accuracy and F1 score of 6 ML models for 3 settings for training data: 1) in original form after conducting feature engineering and standardised on raw data (original), 2) is standardised and normalized and resampled using SMOTE (Processed), 3) after applying PCA on standardised and resampled data (with PCA). We have used 80% data to train the models and 20% for validation. As observed in the result, Random Forest (RF) is the best classifier, followed by Support Vector Machine (SVM) for the original data. However, SVM was the best performer for the pre-processed version. After applying PCA, RF performed slightly better than SVM.

Table 1: Performance comparison of different ML models for the 3-activity classification task

ML Model	Original		Processed		with PCA	
	Acc.	F1	Acc.	F1	Acc.	F1
Logistic Regression	0.902	0.891	0.808	0.808	0.746	0.746
Decision Tree	0.906	0.903	0.899	0.898	0.873	0.872
Random Forest	0.921	0.913	0.968	0.968	0.960	0.960
K-Nearest Neighbors	0.912	0.906	0.895	0.889	0.912	0.908
Support Vector Machine	0.916	0.912	0.981	0.981	0.972	0.972
Neural Network	0.899	0.894	0.974	0.974	0.954	0.954

Table 2: Performance comparison of different ML models for 7-activity classification task

ML Model	Original		Processed		with PCA	
	Acc.	F1	Acc.	F1	Acc.	F1
Logistic Regression	0.684	0.605	0.633	0.624	0.488	0.473
Decision Tree	0.684	0.602	0.781	0.779	0.756	0.750
Random Forest	0.709	0.626	0.949	0.949	0.946	0.946
K-Nearest Neighbors	0.681	0.615	0.873	0.840	0.876	0.847
Support Vector Machine	0.702	0.617	0.956	0.955	0.944	0.941
Neural Network	0.580	0.586	0.907	0.905	0.848	0.845

4.2 Performance evaluation using DL models

In the literature, CNN-LSTM is dominantly used as a DL model for HAR. We have used the raw data for DL classifier. Here we developed and compared four different variants of CNN-LSTM [10, 12, 20, 21], specifically designed to classify sensor-based time-series data. The models operate on fixed-length windows of multi-channel sensor inputs (accelerometer and gyroscope signals) and are processed through convolutional and recurrent layers. The four variants are described as follows:

- **Baseline CNN-LSTM** This comprised two convolutional layers (64 and 128 filters), each followed by max pooling, an LSTM layer with 100 units, and two dense layers (50 and 7 units) with a dropout of 0.5 before classification.
- **CNN-LSTM with Attention** This modifies the baseline by incorporating a self-attention mechanism immediately after the LSTM output. Instead of feeding the entire output sequence directly into dense layers, the model computes a learned weighted sum over the LSTM outputs using a trainable attention layer. The attention vector is then concatenated with the global LSTM output and passed through a dense layer before classification.
- **CNN-LSTM with Cross Attention** Introduces a cross-attention mechanism between two separate CNN-LSTM pipelines. Each input branch processes data from one hand using its own convolutional and recurrent layers. After the LSTM stage, the two streams interact via a cross-attention mechanism, where each stream attends to the other's output sequence. This enables the model to learn contextual relationships and interdependencies between the two hands. The attended outputs are concatenated and passed through shared dense layers to produce the final prediction.
- **CNN-LSTM without Max Pooling** To assess the effect of temporal resolution loss due to pooling, this model removes all max pooling layers from the cross-attention model and uses global pooling instead.

Training was performed using the Adam optimiser with an initial learning rate of 0.001 and the 'categorical_crossentropy' loss function. The batch size was set to 64, and each model was trained for 100 epochs. The dataset was split into training and test sets using an 80:20 ratio, which was used across all models for fair comparison. Again, we did separate experiments for the 3-class and 7-class activity datasets. The Accuracy (A) of all models for data processed with two tolerance values (T=0.01 and T=0.02) results are presented in Table 3.

Table 3: Performance comparison across DL models for the 3 and 7-activity classification tasks.

CNN-LSTM model	3-classes		7-classes	
	A(T=0.01)	A(T=0.02)	A(T=0.01)	A(T=0.02)
Basic	0.944	0.924	0.795	0.726
Attention	0.922	0.925	0.765	0.777
Cross-Attention	0.949	0.941	0.810	0.739
No Max Pooling	0.955	0.951	0.846	0.807

The CNN-LSTM with no max pooling outperforms others. Since T=0.01 had higher data samples, it performed better than T=0.02.

Overall, all these observed results indicate that our dataset is feasible to use for future research or our developed model can be deployed to detect operator activities in the meat processing plant.

5 Conclusion

We presented real-world data for recognising human operator activities in industrial environments. Our dataset includes hand motion data of meat processing plant operators while they work with knives and process meat in real industrial settings. To make the data usable, we released two versions: raw data and feature-engineered data. We conducted extensive experiments using ML and DL models to evaluate the performance of our dataset. Through comprehensive experiments and analysis of the results, we validated the effectiveness and robustness of our dataset in accurately classifying operator activities in industrial environments. Data and sample ML and DL models are available at https://github.com/DigitalInnovationLab/mpp_dataset.

Acknowledgments

The authors would like to acknowledge the valuable support of AMPC Australia, its members, stakeholders, and partners.

GenAI Usage Disclosure

ChatGPT and CoPilot are used to enhance the sentence structure in few places in this paper. Moreover, Gen Ai has been used to generate Python code to train a machine learning model and obtain evaluation results. Gen AI is also used to generate table from spreadsheet that has been used in this paper.

References

[1] Hervé Abdi and Lynne J Williams. 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* 2, 4 (2010), 433–459.
[2] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge Luis Reyes-Ortiz, et al. 2013. A public domain dataset for human activity recognition using smartphones.. In *Esann*, Vol. 3. 3–4.
[3] Abhik Banerjee, Kaneez Fizza, Dimitrios Georgakopoulos, Abdur Rahim Mohammad Forkan, Prem Prakash Jayaraman, and Josip Karabotic Milovac. 2024. Improving the High-Quality Product Consistency in a Digital Manufacturing Environment. *IEEE Transactions on Industrial Informatics* (2024).
[4] Shing Chan, Yuan Hang, Catherine Tong, Aidan Acquah, Abram Schonfeldt, Jonathan Gershuny, and Aiden Doherty. 2024. CAPTURE-24: A large dataset of wrist-worn activity tracker data collected in the wild for human activity recognition. *Scientific Data* 11, 1 (2024), 1135.
[5] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.
[6] M. DALLEL, V. HAVARD, D. BAUDRY, and X. SAVATIER. 2020. InHARD - Industrial Human Action Recognition Dataset in the Context of Industrial Collaborative Robotics. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*. 1–6.
[7] Robin Fays and Rim Slama. 2021. Activity Recognition in Industrial Environment using Two Layers Learning. In *Proceedings of the 2021 3rd International Conference on Video, Signal and Image Processing*. 25–30.
[8] Abdur Rahim Mohammad Forkan, Federico Montori, Dimitrios Georgakopoulos, Prem Prakash Jayaraman, Ali Yavari, and Ahsan Morshed. 2019. An industrial IoT solution for evaluating workers' performance via activity recognition. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 1393–1403.
[9] Francesco Iodice, Elena De Momi, and Arash Ajoudani. 2022. Hri30: An action recognition dataset for industrial human-robot interaction. In *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, 4941–4947.
[10] Mst Alema Khatun, Mohammad Abu Yousuf, Sabbir Ahmed, Md Zia Uddin, Salem A Alyami, Samer Al-Ashhab, Hanan F Akhdar, Asaduzzaman Khan, Akm Azad, and Mohammad Ali Moni. 2022. Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor. *IEEE Journal of Translational Engineering in Health and Medicine* 10 (2022), 1–16.
[11] Xiang TR Kong, Hao Luo, George Q Huang, and Xuan Yang. 2019. Industrial wearable system: the human-centric empowering technology in Industry 4.0. *Journal of Intelligent Manufacturing* 30 (2019), 2853–2869.
[12] Enes Koşar and Billur Barshan. 2023. A new CNN-LSTM architecture for activity recognition employing wearable motion sensor data: Enabling diverse feature extraction. *Engineering Applications of Artificial Intelligence* 124 (2023), 106529.
[13] Xiulong Liu, Jiannong Cao, Yanni Yang, and Shan Jiang. 2018. CPS-based smart warehouse for industry 4.0: A survey of the underlying technologies. *Computers* 7, 1 (2018), 13.
[14] Aleksej Logacjov, Kerstin Bach, Atle Kongsvold, Hilde Bremseth Bårdstu, and Paul Jarle Mork. 2021. HARTH: a human activity recognition dataset for machine learning. *Sensors* 21, 23 (2021), 7853.
[15] A Mastakouris, G Andriosopoulou, D Masouros, P Benardos, G-C Vosniakos, and D Soudris. 2023. Human worker activity recognition in a production floor environment through deep learning. *Journal of Manufacturing Systems* 71 (2023), 115–130.
[16] Vishal Patel, Austin Chesmore, Christopher M Legner, and Santosh Pandey. 2022. Trends in workplace wearable technologies and connected-worker solutions for next-generation occupational safety, health, and productivity. *Advanced Intelligent Systems* 4, 1 (2022), 2100099.
[17] Christopher Reining, Friedrich Niemann, Fernando Moya Rueda, Gernot A Fink, and Michael ten Hompel. 2019. Human activity recognition for production and logistics—a systematic literature review. *Information* 10, 8 (2019), 245.
[18] Wenjin Tao, Ze-Hao Lai, Ming C Leu, and Zhaozheng Yin. 2018. Worker activity recognition in smart manufacturing using imu and semg signals with convolutional neural networks. *Procedia Manufacturing* 26 (2018), 1159–1166.
[19] Zipeng Wang and Jihong Yan. 2024. Multi-sensor fusion based industrial action recognition method under the environment of intelligent manufacturing. *Journal of Manufacturing Systems* 74 (2024), 575–586.
[20] Kun Xia, Jianguang Huang, and Hanyu Wang. 2020. LSTM-CNN architecture for human activity recognition. *Ieee Access* 8 (2020), 56855–56866.
[21] Xiaochun Yin, Zengguang Liu, Deyong Liu, and Xiaojun Ren. 2022. A Novel CNN-based Bi-LSTM parallel model with attention mechanism for human activity recognition with noisy data. *Scientific Reports* 12, 1 (2022), 7878.