

# Continuous Human Activity Recognition in Logistics from Inertial Sensor Data using Temporal Convolutions in CNN

Abbas Shah Syed<sup>1</sup>  
University of Louisville, USA

Zafi Sherhan Syed<sup>2</sup>  
Mehran University, Pakistan

Arez Khalil Memon<sup>3</sup>  
University of Electronic Science and  
Technology of China (UESTC), P. R. China

**Abstract**—Human activity recognition has been an important task for the research community. With the introduction of deep learning architectures, the performance of activity recognition algorithms has improved significantly. However, most of the research in this area has focused on activity recognition for health/assisted living with other applications being given less attention. This paper considers continuous activity recognition in logistics (order picking and packing operations) using a convolutional neural network with temporal convolutions on inertial measurement sensor data from the recently released LARA dataset. Four variants of the popular CNN-IMU are experimented upon and a discussion of the results is provided. The results indicate that temporal convolutions are able to achieve satisfactory performance for some activities (hand center and cart) whereas they perform poorly for the activities of stand and hand up.

**Keywords**—Convolutional Neural Networks; deep learning; Human Activity Recognition (HAR); inertial sensors; LARA dataset

## I. INTRODUCTION

Activity recognition has been an important task for researchers in the field of gaming [1], assisted living [2], sports analysis [3], logistics and other industrial operations [4] and for monitoring of patients diseases, such as Parkinsons to build activity profiles for therapeutic purposes [5]. Eventhough activity recognition, generally speaking, can be performed in a variety of ways [6], [7], [8], inertial sensors have been far by the most popular modality to use for this purpose. This is due to the fact that they are mobile, less cumbersome to wear and cost less than sensing devices for other modalities. Moreover, with their incorporation in phones and smart watches etc, these sensors are usually easily available to the subject for use in activity recognition tasks. This ubiquitous presence combined with ease of data collection has resulted in large datasets being produced which has led to the use of deep learning for various aspects of activity recognition tasks as shown in [9], [10].

As mentioned, the field of activity recognition has been of attention to researchers in various domains, a domain that has not received as much interest is activity recognition in industrial settings. In this paper, data from the LARA dataset [11] is utilized to perform continuous recognition of activities in a logistics scenario using two different convolutional neural network (CNN) architectures, one is a typical convolutional network and the other is a modified version of the parallel CNN architecture called CNN-IMU suggested in [12] which performs convolutions in the temporal domain. The experiments

indicated that the parallel CNN architecture performed better than the considered typical CNN. The rest of the paper is organized as follows: section II discusses previous work carried out for activity recognition, section III presents an overview of the dataset used, section IV presents the methodology of the paper; the pre-processing steps and the discussion of the networks used in the experiments, section V discusses the results for the current work with a conclusion being provided in section VI and future directions in section VII.

## II. LITERATURE REVIEW

Activity recognition has been at the forefront of pervasive computing research and the development of cyber physical systems as this has enormous societal and economical impact potential [13]. This section covers previous work in the direction of activity recognition using utilizing inertial measurement sensors (IMUs).

Industrial activity recognition using IMUs has been targeted by multiple research works for varying applications, these include, wood shops [14], construction [15], assembly line [16], process optimization [17]. An early work using deep learning methods for activity recognition for industry was suggested by [18] on the Skoda dataset [19]. Their network consists of one convolutional layer, one pooling layer, two hidden layers and one softmax layer for classification. Moreover, the convolutional layer contains several convolutional blocks in parallel with partial weight sharing for the three axes of accelerometer sensor values. The pooling layer also pools convolutional blocks sharing their weights separately before the outputs are passed on to the later layers.

An interesting approach for segmenting different types of activities for health risk assessment in an order picking process is presented in [20] who utilize angles of human body joints from 17 IMUs placed on a workers body performing the order picking activity. Joint angles between body limbs are computed using an extended Kalman filter [21] and these are used to segment the sub-activities within the picking process. Risk assessment is performed using the rapid entire body assessment (REBA) standard [22]. The authors in [23] propose using accelerometer and gyroscope signal data to perform activity recognition for worker performance assessment in the meat industry. After extracting segments, they compute several features from both sensors and test the performance of multiple classifiers for determining output activity. Following from their

work in [24], the authors in [25] combine video and inertial measurement sensor data to determine grabbing actions in a picking process. They do this by extracting various time and frequency domain features from the IMU data as well as colour and descriptor features from the video data and then passing them on to three machine learning classifiers for prediction purposes.

The approach suggested in [12] makes use of CNNs on the data provided in [17]. They propose a CNN architecture called CNN-IMU based on the CNN proposed in [26]. Their network uses four CNN layers with parallel CNN blocks sharing their weights for each IMU, two pooling operations, fully connected layer before they are combined using a fully connected layer followed by a softmax for classification. The motivation to spread the network instead of making it deeper is that it becomes more descriptive. The authors in [27] make use of accelerometer sensor data in the dataset in [28] and CNNs to differentiate between activities in the industry. They achieve their best results when using raw signal data and sending it to the CNN for classification. The authors in [29] use the CNN-IMU on three different datasets of activity recognition, two consisting of various different activities in daily life and one of activities from a logistic scenario of order picking. Their experiments involved comparing two different CNN architectures, one a CNN-IMU and the other a baseline CNN with the same layerwise make up as the CNN-IMU. For the logistics scenario, their result indicates that the CNN-IMU outperforms the typical CNN in nearly all experiments. Their results indicate to the effectiveness of using wider networks consisting of parallel layers instead of using deeper ones. CNNs have also been used with other modalities for industrial processes, in [30] semantic representations have been used for activity recognition in a picking setup. Motion capture data from the MoCAP dataset is used with the CNN architecture described in [31]. The authors in [11] use the Logistics Activity Recognition Challenge (LARA) dataset provided by them to determine activities in a logistics scenario. They carry this out by using a modified version of the t-CNN in [26] which consists of four convolutional layers, two fully connected layers followed by two separate softmax and sigmoid layers to determine the sub-activity being performed and the attribute from an activity attribute list on motion capture data from the dataset.

It can be observed from the literature review that convolutional neural networks have proved to be very useful for performing activity recognition in industrial scenarios. This paper utilizes a convolutional neural network for performing continuous activity recognition for logistics using inertial sensor data from the LARA dataset. This paper compares the performance of a modified version of the CNN-IMU network presented in [26] and used in [11] to a typical CNN. The CNN utilizes convolutions in the temporal axis to extract important features from time series sensor data and is well suited for use with inertial measurement sensor signals.

### III. DATASET

The LARA dataset provides data of multiple modalities from recordings in a logistics scenario. Video recordings, Motion Capture data and data from inertial measurement units is recorded from 14 people in the dataset. Each of

the participants is asked to perform three tasks which are common in logistics operations, two of these are picking tasks and the third is packing. Motion capture data was captured using a Optical Marker-based Motion Capture (OMoCap) system which resulted in markers for the movements of the participants, moreover, several IMUs were used to record the movement patterns too along with RGB videos of the activities being performed. The total duration of the recorded data is 758 minutes which has been annotated in two ways, first, an annotation is provided for each intra-activity that comprises the picking and packing tasks and second, binary semantic representations of a different type of representation for the picking and packing tasks. The first representation represents the activities in terms of eight intra-activities and is used in this work and the second type of representation as attributes to describe the task recordings. The eight annotations are standing, walking, cart (participant is walking with the cart), handling upwards (participant has atleast one hand raised upward to shoulder height), handling centered (participant can handle things without bending, lifting their arms or needing to kneel), handling downwards (participant has hands below his knees while kneeling or otherwise), synchronization (waving motion before each recording) and a set of samples which were unrecognizable by the annotators and have been marked as *None*.

This dataset provides the opportunity to develop algorithms for both the picking and packing operations in logistics by containing recordings of multiple modalities to researchers. From these modalities, this paper focuses on the data from the IMUs collected in these experiments. Three types of IMUs were used in the trials with 14 people in total performing the said tasks with data being collected from five points on the body, both the arms, legs and the chest/mid-body. The sampling frequency for the IMUs is 100 Hz. A summary of the recordings present in the dataset are given in Table I. Readers interested in more detail are referred to [11].

TABLE I. SUMMARY OF IMU MEASUREMENTS IN THE LARA DATASET

Subject ID	Gender	Age	Scenario 1	Scenario 2	Scenario 3
S07	M	23	2	13	14
S08	F	51	2	13	14
S09	M	35	2	14	13
S10	M	49	2	13	12
S11	F	47	2	12	0
S12	F	23	0	6	14
S13	F	25	2	14	14
S14	M	54	2	14	14
<b>Total</b>			14	99	95

### IV. METHODOLOGY

To perform continuous monitoring of activities in logistics, we use a two step process. Segments are first extracted from the IMU sensor data for each trial which are then passed to the CNNs to test their performance. For the first two experiments, segmentation is performed for all IMUs together whereas for the last two experiments data segmentation takes place for each of the five IMUs individually. These are then fed to the CNN networks as inputs.

#### A. Preprocessing Stage

Windows of 100 samples are extracted from the recording for each sensor and position with a step size of 25 samples

(75% overlap) for successive windowed segments. An overlap is used to ensure that enough samples are generated to develop a large enough dataset for training of deep learning networks. Furthermore, since the annotations in the dataset are present on a sample by sample basis for each value of each sensor, an extracted segment is assigned a *segment annotation* by majority voting of the annotations of its samples. We then use the segment annotations as the appropriate annotation labels, a similar approach has been used in [29]. Once segments have been extracted from all the trials for all subjects, the segments belonging to the *synchronization* and *None* class are removed. The rest of the segments are used for classification with the convolutional neural network.

### B. Classification

For classification, we make use of a modified version of the convolutional neural network described in [12] named CNN-IMU. They propose a network which utilizes four convolutional layers, two max pooling layers, fully connected layer and a softmax layer for determining the output class. Each of the convolutional layers have multiple parallel convolutional blocks sharing their weights which perform convolution operations along the time axis. The number of convolutional blocks in each layer depends on the number of IMUs present in the data, one block for each IMU. Input to the network is provided as windowed segments of IMU sensor signal recordings over the temporal domain. The output of each of these parallel blocks and pooling layers is passed to a fully connected layer individually for computing an intermediate feature representation. These representations are then combined using a fully connected layer before being passed to the softmax layer for classification. Dropout was applied to the fully connected layers apart from the softmax layer.

This CNN-IMU network is used in two different variants in this work based on the results of [29] and [11]. The first network follows the network construction as described in [12] which includes the max pooling operations of the network. The second variant of the network skips the pooling operations as was used in [11] which were found to affect network performance negatively in [31]. Both these variants require that data from the IMUs be segmented individually as separate parallel inputs. Furthermore, to compare the performance of the two CNN-IMU architectures to typical CNNs utilizing temporal convolutions, typical CNNs that do not make use of parallel convolutional blocks for different IMUs but consist of the same layer-wise structure are used to perform classification as well. These networks require that the segments be extracted for all five IMUs as one frame/segment. The details of the networks are provided in Table II and the architectures are illustrated in Fig. 1. The pooling layers have been shown with a dotted border to indicate the absence of these operations in the network variations considered in this research work.

The data from the sensors was split in to train, validation and test sets and the network was trained using an Adam Optimizer with cross entropy as the loss function. A learning rate of  $1 \times 10^{-6}$  was used along with a batch size of 400. Moreover, training was performed for 12 epochs with early stopping utilized to retain the best model.

TABLE II. DETAILS OF CNN ARCHITECTURES

CNN Architecture	Number of Layers		
	Convolutional	Pooling	FC + Softmax
Typical CNN-1	4	2	3
Typical CNN-2	4	0	3
CNN-IMU-1	4 (Five blocks/ layer)	2	3
CNN-IMU-2	4 (Five blocks / layer)	0	3

## V. EXPERIMENTATION, RESULTS AND DISCUSSION

In order to test the efficacy of the four CNNs considered in this work, we perform experiments for each of the four networks individually and report on the results obtained. The results for each experiment are reported in terms of the Precision, Recall and F1 score for each activity class.

### A. Experiment with Typical CNN-1

In this experiment we used the typical CNN-1 architecture that consists of 4 convolution layers, 2 max pooling layers, 2 fully connected layers and one softmax layer. The input to this network were segments/ frames consisting of the combined information from all the IMUs together. The results for the classification are provided in Table III. It can be observed from the table that the CNN has performed poorly for the activities *Hand up* and *Stand*. Moreover, as can be observed, the best performing activities were *Cart* and *Hand Center*.

TABLE III. SUMMARY OF RESULTS: TYPICAL CNN-1 (WITH MAX-POOLING)

Activity	Precision (%)	Recall (%)	F1 (%)
Stand	38.73	61.96	47.66
Walk	71.27	70.76	71.01
Cart	82.44	86.31	84.33
Hand up	45.61	79.48	57.96
Hand Center	91.87	80.75	85.95
Hand Down	71.93	76.44	74.12

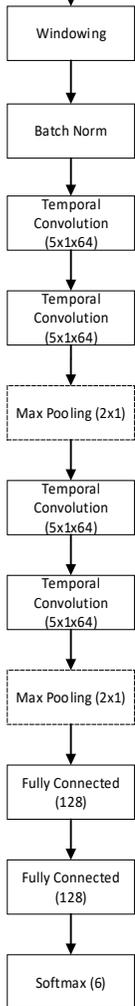
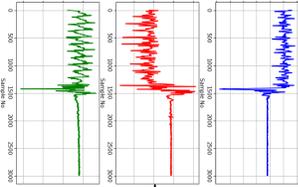
### B. Experiment with Typical CNN-2

For the second experiment, we use the typical CNN-2 architecture consisting of 4 convolution layers, 2 fully connected layers and one softmax layer. The input to this network too were segments/ frames consisting of the combined information from all the IMUs together. The results of the classification are presented in Table IV. It can be observed that the performance for this network is very similar to the *Typical CNN-1 network* of experiment 1 which used maxpooling layers, there has been some degradation in performance for some activities. In this experiment too, the network was able to best recognize the activities of *Cart* and *Hand Center* while poor performance was observed for the activities of *Stand* and *Hand up*.

TABLE IV. SUMMARY OF RESULTS: TYPICAL CNN-2 (WITHOUT MAX-POOLING)

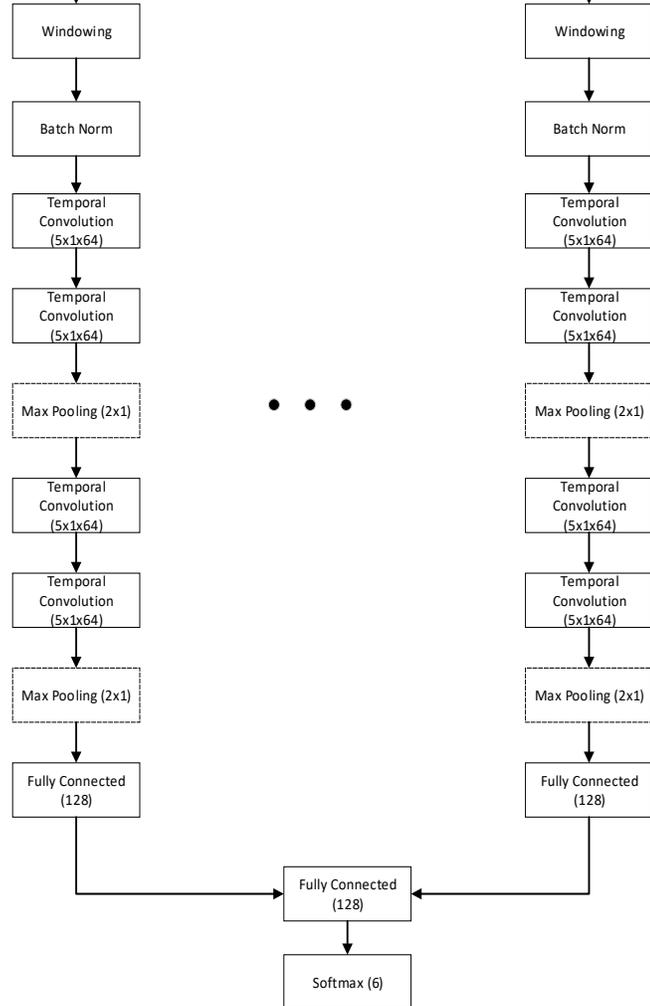
Activity	Precision (%)	Recall (%)	F1 (%)
Stand	39.06	47.95	43.05
Walk	57.04	81.01	66.94
Cart	80.38	85.42	82.83
Hand up	34.99	79.03	48.51
Hand Center	92.21	78.99	85.09
Hand Down	71.23	77.04	74.02

All IMU data segmented together



(a)

Data from IMUs segmented individually



(b)

Fig. 1. CNN Architectures used in this Work (a) Typical CNN (b) CNN-IMU.

### C. Experiment with CNN-IMU-1

This experiment involved the usage of the CNN-IMU-1 network to determine activity classes. This architecture consists of four convolutional layers with five convolutional blocks, one for each IMU data and these blocks share their weights. The IMU segments are individually fed to the blocks resulting in the classification scores shown in Table V. This network has also been able to recognize the activities of *Hand Center* and *Cart* well but the activity *Stand* is still the worst performing activity among the six activity classes being considered.

TABLE V. SUMMARY OF RESULTS: CNN-IMU-1 (WITH MAX-POOLING)

Activity	Precision (%)	Recall (%)	F1 (%)
Stand	27.68	66.85	39.15
Walk	69.42	67.03	68.20
Cart	87.95	79.67	83.61
Hand up	43.62	60.18	50.58
Hand Center	90.34	80.04	84.88
Hand Down	71.93	75.68	73.76

### D. Experiment with CNN-IMU-2

In this experiment we used the CNN-IMU-2 architecture which consists of four convolution layers, two fully connected layers and one softmax layer. The input to this network were segments/ frames consisting of the individual IMUs. The classification results are presented in Table VI. The omission of the pooling layer has impacted network performance positively as was observed in other works. Similar to the previous cases, this network produces the best results for the activities *Cart* and *Hand Center*.

TABLE VI. SUMMARY OF RESULTS: CNN-IMU-2 (WITHOUT MAX-POOLING)

Activity	Precision (%)	Recall (%)	F1 (%)
Stand	40.07	47.11	43.31
Walk	49.08	82.73	61.61
Cart	81.79	83.40	82.59
Hand up	26.04	73.36	38.43
Hand Center	91.87	77.88	84.29
Hand Down	71.46	75.37	73.37

Following from the experiments conducted in this work with the considered CNN architectures, the most suitable network for continuous activity recognition from inertial sensor data was found to be the Typical CNN-1 architecture which involves pooling operations. The best scores are achieved for the activities *Cart* and *Hand Cent* whereas the worst scores have been produced for the activities *Hand Up* and *Stand*, this was the case for all the networks considered in this research work. The F1 scores for each of the networks for the six activities are listed in Table VII.

## VI. CONCLUSION

This paper explores the usage of temporal convolutions in a CNN for the problem of continuous activity recognition in a logistics scenario using inertial measurement sensor data. Data from the LARA dataset which consists of video, OMOCap and IMU signal recordings from seven different people performing three different tasks concerning picking and packing has been used in this work. To accomplish the aims of this work, four

TABLE VII. SUMMARY OF RESULTS (F1 SCORE [%] FOR EACH CONSIDERED NETWORK)

Activity	F1 (%)			
	Typical CNN-1	Typical CNN-2	CNN-IMU-1	CNN-IMU-2
Stand	47.66484	43.05043	39.14759	43.30519
Walk	71.01347	66.94491	68.20405	61.60714
Cart	84.32787	82.82695	83.60756	82.589
Hand up	57.95574	48.50575	50.57692	38.43329
Hand Center	85.95271	85.0872	84.8763	84.29462
Hand Down	74.11908	74.01961	73.76058	73.36562

CNN architectures, have been tested which take windowed segments of IMU recordings. From the experiments conducted, the typical CNN-1 architecture involving pooling operations was found to be the best performing model. High scores were achieved for the activities *Hand Center* and *Cart*; however, scores for the activity *Stand* and *Hand Up* weren't satisfactory. While satisfactory performance was achieved for the former activities, the performance of the considered networks for the latter activities was poor. Therefore, modifications need to be made for improvement of the network for such activities.

## VII. FUTURE WORK

This work presents experimental work for the continuous recognition of activities for logistics using the LARA dataset. In this paper, only CNN architectures have been considered, for future attempts at this task, other deep learning architectures could be considered to improve activity recognition such as Recurrent Neural Networks with attention, etc. Moreover, sensor fusion could also be used, especially OMOCap data from LARA dataset could be fused with IMU data and used with various deep learning networks to check for performance. Video data could also be combined to create a multimodal solution for activity recognition as suggested in [25]. Dependable activity recognition systems will help in the optimization of industrial processes as well as be used for health assessment purposes.

## REFERENCES

- [1] N. Jablonsky, S. McKenzie, S. Bangay, and T. Wilkin, "Evaluating sensor placement and modality for activity recognition in active games," in *Proceedings of the Australasian Computer Science Week Multiconference*, 2017, pp. 1–8.
- [2] J. Rafferty, C. D. Nugent, J. Liu, and L. Chen, "From activity recognition to intention recognition for assisted living within smart homes," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 3, pp. 368–379, 2017.
- [3] Y.-L. Hsu, S.-C. Yang, H.-C. Chang, and H.-C. Lai, "Human daily and sport activity recognition using a wearable inertial sensor network," *IEEE Access*, vol. 6, pp. 31 715–31 728, 2018.
- [4] C. Reining, F. Niemann, F. Moya Rueda, G. A. Fink, and M. ten Hompel, "Human activity recognition for production and logistics—a systematic literature review," *Information*, vol. 10, no. 8, p. 245, 2019.
- [5] W.-Y. Cheng, A. Scotland, F. Lipsmeier, T. Kilchenmann, L. Jin, J. Schjodt-Eriksen, D. Wolf, Y.-P. Zhang-Schaerer, I. F. Garcia, J. Siebourg-Polster *et al.*, "Human activity recognition from sensor-based large-scale continuous monitoring of parkinson's disease patients," in *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 2017, pp. 249–250.
- [6] S.-R. Ke, H. L. U. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, and K.-H. Choi, "A review on video-based human activity recognition," *Computers*, vol. 2, no. 2, pp. 88–131, 2013.

- [7] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, pp. 1–33, 2014.
- [8] X. Luo, Q. Guan, H. Tan, L. Gao, Z. Wang, and X. Luo, "Simultaneous indoor tracking and activity recognition using pyroelectric infrared sensors," *Sensors*, vol. 17, no. 8, p. 1738, 2017.
- [9] C. Xu, D. Chai, J. He, X. Zhang, and S. Duan, "Innohar: a deep neural network for complex human activity recognition," *Ieee Access*, vol. 7, pp. 9893–9902, 2019.
- [10] G. L. Santos, P. T. Endo, K. H. d. C. Monteiro, E. d. S. Rocha, I. Silva, and T. Lynn, "Accelerometer-based human fall detection using convolutional neural networks," *Sensors*, vol. 19, no. 7, p. 1644, 2019.
- [11] F. Niemann, C. Reining, F. M. Rueda, N. R. Nair, J. A. Steffens, G. A. Fink, and M. t. Hompel, "Lara: Creating a dataset for human activity recognition in logistics using semantic attributes," *Sensors*, vol. 20, no. 15, p. 4083, 2020.
- [12] R. Grzeszick, J. M. Lenk, F. M. Rueda, G. A. Fink, S. Feldhorst, and M. ten Hompel, "Deep neural network based human activity recognition for the order picking process," in *Proceedings of the 4th international Workshop on Sensor-based Activity Recognition and Interaction*, 2017, pp. 1–6.
- [13] P. Nikolov, O. Boumbarov, A. Manolova, K. Tonchev, and V. Poulkov, "Skeleton-based human activity recognition by spatio-temporal representation and convolutional neural networks with application to cyber physical systems with human in the loop," in *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, 2018, pp. 1–5.
- [14] J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 10, pp. 1553–1567, 2006.
- [15] J. Zhao and E. Obonyo, "Towards a data-driven approach to injury prevention in construction," in *Workshop of the European Group for Intelligent Computing in Engineering*. Springer, 2018, pp. 385–411.
- [16] H. Koskimäki, V. Huikari, P. Siirtola, and J. Rönning, "Behavior modeling in industrial assembly lines using a wrist-worn inertial measurement unit," *Journal of Ambient Intelligence and Humanized Computing*, vol. 4, no. 2, pp. 187–194, 2013.
- [17] S. Feldhorst, M. Masoudenijad, M. ten Hompel, and G. A. Fink, "Motion classification for analyzing the order picking process using mobile sensors," in *Proc. Int. Conf. Pattern Recognition Applications and Methods*, 2016, pp. 706–713.
- [18] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, "Convolutional neural networks for human activity recognition using mobile sensors," in *6th International Conference on Mobile Computing, Applications and Services*. IEEE, 2014, pp. 197–205.
- [19] P. Zappi, C. Lombriser, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Tröster, "Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection," in *European Conference on Wireless Sensor Networks*. Springer, 2008, pp. 17–33.
- [20] T. Hara, Y. Li, J. Ota, and T. Arai, "Automatic risk assessment integrated with activity segmentation in the order picking process to support health management," *CIRP Annals*, 2020.
- [21] C. Jakob, P. Kugler, F. Hebenstreit, S. Reinfelder, U. Jensen, D. Schuldhau, M. Lochmann, and B. M. Eskofier, "Estimation of the knee flexion-extension angle during dynamic sport motions using body-worn inertial sensors," in *Proceedings of the 8th International Conference on Body Area Networks*, 2013, pp. 289–295.
- [22] S. Hignett and L. McAtamney, "Rapid entire body assessment (reba)," *Applied ergonomics*, vol. 31, no. 2, pp. 201–205, 2000.
- [23] A. R. M. Forkan, F. Montori, D. Georgakopoulos, P. P. Jayaraman, A. Yavari, and A. Morshed, "An industrial iot solution for evaluating workers' performance via activity recognition," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019, pp. 1393–1403.
- [24] A. Diete, L. Weiland, T. Szytler, and H. Stuckenschmidt, "Exploring a multi-sensor picking process in the future warehouse," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 2016, pp. 1755–1758.
- [25] A. Diete, T. Szytler, L. Weiland, and H. Stuckenschmidt, "Recognizing grabbing actions from inertial and video sensor data in a warehouse scenario," in *FNC/MobiSPC*, 2017, pp. 16–23.
- [26] J. Yang, M. N. Nguyen, P. P. San, X. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *Ijcai*, vol. 15. Citeseer, 2015, pp. 3995–4001.
- [27] X. Zheng, M. Wang, and J. Ordieres-Meré, "Comparison of data preprocessing approaches for applying deep learning to human activity recognition in the context of industry 4.0," *Sensors*, vol. 18, no. 7, p. 2146, 2018.
- [28] T. Szytler and H. Stuckenschmidt, "On-body localization of wearable devices: An investigation of position-aware activity recognition," in *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2016, pp. 1–9.
- [29] F. Moya Rueda, R. Grzeszick, G. A. Fink, S. Feldhorst, and M. Ten Hompel, "Convolutional neural networks for human activity recognition using body-worn sensors," in *Informatics*, vol. 5, no. 2. Multidisciplinary Digital Publishing Institute, 2018, p. 26.
- [30] C. Reining, M. Schlangen, L. Hissmann, M. ten Hompel, F. Moya, and G. A. Fink, "Attribute representation for human activity recognition of manual order picking activities," in *Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction*, 2018, pp. 1–10.
- [31] F. M. Rueda and G. A. Fink, "Learning attribute representation for human activity recognition," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 523–528.