TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)
31 Oct - 3 Nov 2023. Chiang Mai, Thailand

Human Activity Recognition in Logistics Using Wearable Sensors and Deep Residual Network

Sakorn Mekruksavanich[1], Datchakorn Tancharoen[2] and Anuchit Jitpattanakul[3]
[1]*Department of Computer Engineering, School of Information and Communication Technology*
University of Phayao, Phayao, Thailand
sakorn.me@up.ac.th
[2]*Faculty of Engineering and Technology, Panyapiwat Institute of Management*, Nonthaburi, Thailand
datchakorntan@pim.ac.th
[3]*Intelligent and Nonlinear Dynamic Innovations Research Center, Department of Mathematics*
Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
anuchit.j@sci.kmutnb.ac.th

***Abstract*—Human action identification is a practical area of study with broad applicability in various domains, such as medical care, sport science, and manufacturing management. In logistics, it is essential to identify and examine individual actions, enabling machines to perceive and comprehend human motions for non-verbal interaction. This study specifically focuses on efficiently classifying working activities in the logistics industry using wearable sensors, particularly in the context of human activity recognition. To achieve the research objective, a deep residual neural network was introduced, integrating convolutional layers, shortcut connections, and aggregated transformation for human activity recognition in logistics. The authors evaluated the effectiveness of their proposed deep learning model using the publicly accessible LARa dataset. The LARa dataset comprises a diverse range of human actions in the logistics domain, including standing, walking, cart handling, and synchronization. The details of activity were captured using wearable sensors affixed to different anatomical sites of the study participants. The experimental findings indicate that the model achieved a maximum F-measure of 85.30%.**

***Keywords*—deep learning, logistics, human activity recognition, wearable sensors**

I. Introduction

The identification of autonomous employee actions presents an opportunity to improve worker effectiveness in terms of efficiency and security while also enhancing project management capabilities. This procedure provides valuable insights into understanding activity levels and identifying factors that can aid in project decision-making. These insights prove helpful in various project management responsibilities, including adapting project schedules, managing resources, and exercising construction-related authority [1].

The increasing prevalence of automated processes in manufacturing and logistics, along with the growing complexity of manual procedures, has led to greater interaction between humans and machines. Effective collaboration relies on interaction, encompassing both verbal and nonverbal connections. Despite mistakenly assuming that the human element is solely determinative in planning and modeling simulations, it remains a critical factor in managing material systems. To successfully execute a data-centric simulation that considers the stochastic locomotion patterns of individuals, the inclusion of temporal data is crucial. Instruments capable of perceiving and understanding human motion play a vital role in enabling non-verbal interaction. One potential approach to achieving this goal is sensor-based human activity recognition (S-HAR) [2], [3].

On-body electronic devices are equipped with three-dimensional inertial sensors that can monitor various types of forces along three distinct axes. These sensors include accelerometers, gyroscopes, and magnetometers. Some of these devices may also incorporate sensors to track vital signs of human beings, such as humidity and pulse rates [4]. Traditional statistical techniques have been used in human activity recognition (HAR) to recognize patterns [5]. These techniques employ a sliding-window approach to segment signal sequences, extract relevant features from these segments, and then use a classifier to assign specific action labels. However, in recent years, deep architectures have proven effective in handling multichannel time-series recognition of human actions, including sports and everyday life scenarios [6]–[9]. Deep architectures integrate feature extraction and classification in an end-to-end manner, which sets them apart from conventional methods [10], [11]. These architectural models are capable of learning non-linear and temporal relationships that exist within complex and dynamic human actions. They achieve this by directly learning from raw inertial data. Furthermore, these models exhibit greater discriminatory power compared to manually created classes of human actions. They also demonstrate invariance to distortions and temporal variations [12].

This study presents a logistic approach to recognize human behaviors using wearable sensors and leverages pre-existing deep learning models. Specifically, a deep residual network is proposed to improve the efficiency of recognition in classifying logistical operations. To evaluate the effectiveness of deep learning models in identifying efficiency, the LARa dataset is utilized. This publicly available dataset allows for comparisons and includes a variety of human interactions in logistics, such as standing, walking, cart handling, and synchronization. The assessment of outcomes and evaluation of model efficacy are performed using metrics such as accuracy, precision, recall, and F-measure.

The following sections of this piece are organized as follows: Section II offers an overview of the latest research relevant to the topic. Section III provides a comprehensive account of the ResNeXt model that has been proposed. Section IV presents the findings obtained from the experiment.

Lastly, Section V concludes with a summary of the study's findings and identifies potential areas for future research that may present significant challenges.

II. Related Works

A. Deep Learning Models in HAR research

Deep learning networks are derived from artificial neural networks (ANNs). Unlike conventional ANNs, which have a shallow architecture with limited hidden layers, deep neural networks (DNNs) have a greater number of layers. This increased layer depth in DNNs allows them to learn from large datasets more effectively. In fact, DNNs are often used as dense layers in various other deep learning models. When it comes to convolutional neural networks (CNNs), it is common to include multiple dense layers after the convolutional layers [13], [14].

CNN is built upon three fundamental concepts: sparse connections, parameter sharing, and equivariant representations [15]. In CNNs, subsequent layers often include pooling and fully connected layers that handle regression or classification tasks. CNNs have shown proficiency in extracting features from signals and have achieved promising results in tasks such as image classification, voice recognition, and text evaluation. When it comes to time series classification, particularly in HAR, the utilization of CNNs offers two notable advantages over alternative models: local dependency and scale invariance. Local dependency refers to the likelihood of associated signals being close to each other in the context of HAR. Scale invariance, on the other hand, refers to the property of remaining consistent across different rates or frequencies.

Recurrent Neural Networks (RNNs) are widely used in speech recognition and natural language processing. This is because RNNs can effectively capture the temporal correlations between neurons, which is crucial for these tasks. It is common to combine RNNs with long-short term memory (LSTM) cells, where LSTM acts as the memory unit during gradient descent. However, there have been limited studies that utilize RNNs for HAR tasks [16]–[19]. These studies primarily focus on optimizing learning efficiency and minimizing resource utilization in HAR applications.

The hybrid model refers to a composite approach that combines different deep models [20]–[22]. One prominent hybrid model is the fusion of CNNs and LSTM models, which has shown promising results [23], [24]. Several approaches have been proposed to integrate CNN and LSTM architectures. In a referenced study [25], it was demonstrated that using recurrent dense layers in conjunction with convolutional neural networks is more effective than using solely dense layers in the same network architecture. Similar findings were also observed in another study [26]. CNNs excel at capturing spatial relationships, while RNNs leverage temporal relationships. By integrating CNNs and RNNs, the capacity to identify diverse activities with distinct temporal durations and signal distributions can be improved. This integration holds potential for enhancing activity recognition in various applications.

B. Logistic Activity Recognition

Several research studies have focused on using Inertial Measurement Units (IMUs) to recognize manufacturing operations in various industries, including timber businesses [27], construction [28], assembly lines [29], and workflow optimization [30]. For instance, a study conducted by Zeng et al. [12] utilized deep learning techniques to recognize activities in the industrial sector, using the Skoda dataset [31]. The network architecture employed in the study consisted of a single convolutional layer, a pooling layer, two hidden layers, and a softmax layer for identification purposes. The convolutional layer incorporated multiple convolutional blocks that operated in parallel, with partial weight distribution specifically designed for the tri-axial accelerometer sensor data. The pooling layer pooled the outputs of these convolutional blocks with distinguished weight distribution before passing them on to subsequent layers.

In another study, Yang et al. [32] utilized the Logistics Activity Recognition Challenge (LARa) dataset, which they curated, to identify different operations within a logistics context. They achieved this using a modified version of the t-CNN model. The modified model comprised four convolutional layers, two fully connected layers, and two distinct softmax and sigmoid layers. These layers were employed to identify the sub-activity being executed and extract relevant characteristics from the set of action characteristics, respectively. The input data for this process was derived from recorded movement information contained within the dataset.

The literature review highlights the significant utility of CNNs in identifying actions within manufacturing contexts. Building upon this, the present study utilizes a CNN to perform uninterrupted activity identification specifically for logistics. Inertial sensor data from the LARa dataset is utilized as the input for the CNN model.

III. Sensor-based HAR Framework

In this study, a sensor-based HAR framework is utilized. The framework consists of four key stages: data acquisition, data pre-processing, data generation, and model training and evaluation. These stages are illustrated in Fig. 1.

A. LARa Dataset

The LARa dataset [33] provides freely accessible multimodal data from logistics situation recordings. This dataset includes video camera footage, motion-capturing information, and data from inertial measurement units (IMUs) collected from a sample of 14 individuals. The study focuses on routine duties commonly performed in logistics processes. Each individual is tasked with completing three assignments, consisting of two picking assignments and one packing assignment. To gather movement data, an Optical Marker-based Motion Capture (OMoCap) system was utilized, which tracked the motions of individuals. In addition, several IMUs were used to capture patterns of movement, and RGB videos were recorded to document the actions performed. The dataset encompasses a total duration of 758 minutes and has been annotated in two distinct forms. The first form of annotation includes annotations for each intra-activity, covering both picking and packing assignments. The second form of annotation involves binary semantic illustrations that

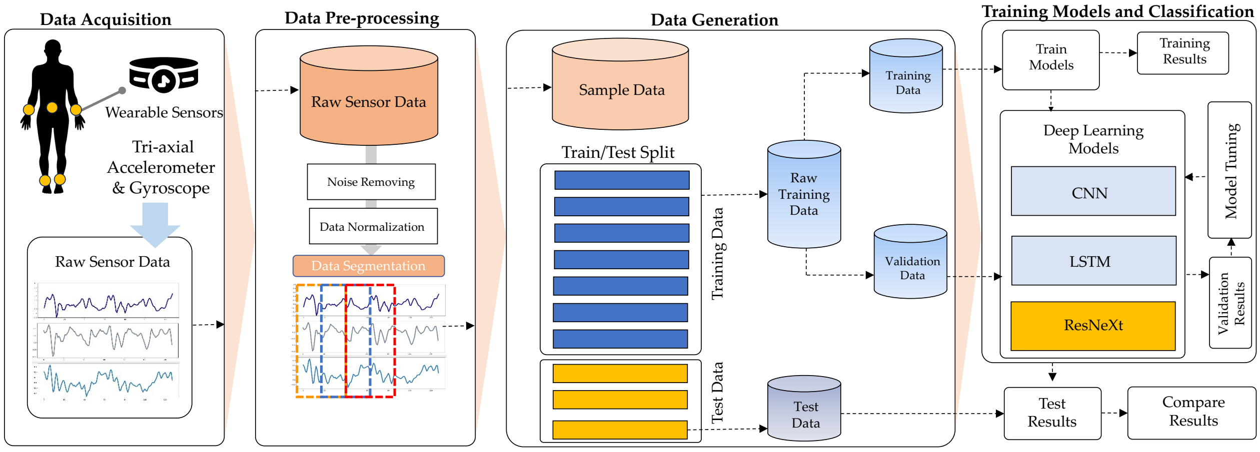

Fig. 1. The HAR framework based on a wearable sensors used in this work.

are specific to the picking and packing assignments. These annotations provide additional context to the dataset.

For this study, the proposed model exclusively employs IMU sensor data sourced from the LARa dataset for training and testing purposes.

B. Data Pre-processing

The pre-processing of the raw sensor data involved two main steps: noise removal and data normalization. Once the data underwent pre-processing, it was divided into segments using a sliding window approach. The length of each segment was fixed at 200 samples, with an overlapping ratio of 50%. This overlapping ratio ensured the generation of sufficient samples to create a comprehensive dataset for training deep learning networks.

Since the dataset contains annotations for each sensor value on a per-sample basis, a segment that was extracted obtained a segment annotation through majority voting of its constituent samples' annotations. These segment annotations were used as appropriate labels for the segments. After extracting segments from each experiment across all participants, segments belonging to the none category were discarded. The remaining segments were employed for categorization using the proposed model.

C. The ResNeXt model

In theory, increasing the number of layers in a deep learning network should improve its effectiveness. However, in practice, it has been observed that an excessive number of layers can lead to challenges such as vanishing or exploding gradients. These issues can negatively impact the network's ability to accurately recognize patterns.

To address this challenge, the ResNeXt architecture offers a solution that enhances accuracy without increasing parameter complexity or reducing the number of hyperparameters. This is achieved through the architecture of the sub-modules. This study demonstrates that by employing parallel stacking of blocks with identical topology, the model's accuracy can be improved without a significant increase in the parameter count. This approach replaces the three-layer convolution block of the original ResNet [34], while maintaining identical topology. As a result, the hyperparameters are correspondingly reduced, making the model more transferable to different tasks.

To enhance more effectiveness of the ResNeXt network model and identify the most distinctive features, an attention mechanism has been incorporated. This mechanism allows for investigating the interdependencies among features. The use of attention mechanisms is a widely adopted technique in deep learning, finding applications in various domains such as natural language processing, image recognition, and speech recognition. By assigning higher weights to significant features, the attention mechanism reduces the network's parameters while enhancing the discriminative power of these features. This approach improves the model's efficiency by focusing on the most relevant information.

The ResNeXt architecture is a comprehensive deep-learning model that incorporates convolutional blocks with multi-kernel residual blocks following the deep residual structure. Fig. 2 provides an illustration of the overall conceptualization of the proposed model.

ResNeXt employs Convolutional Blocks (ConvB) to extract low-level features from raw sensor data. The ConvB model consists of four layers, as shown in Fig. 2. The techniques utilized in this study include Conv1D, BN, ELU, and MP. Conv1D employs trainable convolutional kernels to extract specific features, generating distinct feature maps. The BN layer is chosen to expedite and enhance the training process. The ELU layer is incorporated to amplify the model's expressive power. The MP layer enables compression of the feature map while preserving its essential elements. The Multi-Kernel Blocks (MK) are composed of three parts, utilizing convolutional kernels of varying sizes: 1×3, 1×5, and 1×7. To reduce system complexity and parameter count, the suggested network employs 1×1 convolutions in each component.

The detection block utilized the Global Average Pooling (GAP) technique and flattened layers to convert the mean of each feature map into a one-dimensional vector. The output of the fully connected layer was subjected to Softmax operation to obtain probabilistic inference. To calculate the network's losses, the cross-entropy loss operation, commonly used in classification tasks, was applied.

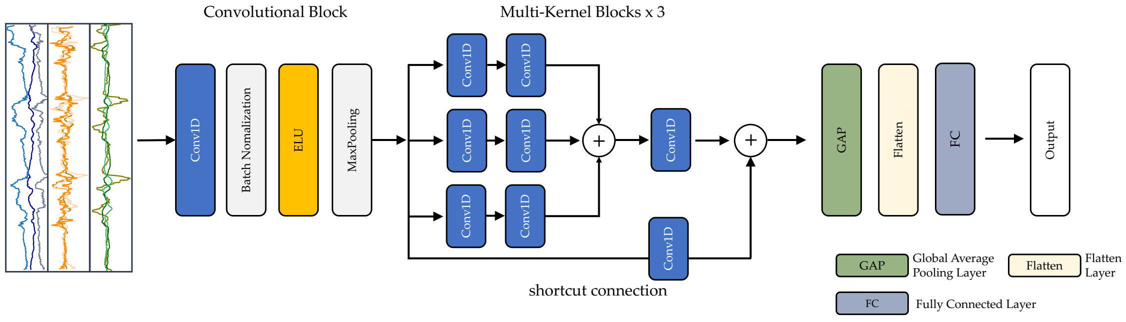

Fig. 2. The ResNeXt architecture used in this work.

D. *Performance Measurement Criteria*

To evaluate the effectiveness of the proposed deep learning model, four widely recognized assessment indicators – accuracy, precision, recall, and F-measure – were calculated using a 5-fold cross-validation approach. The computations for these indicators are provided below:

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{1}$$

$$Precision = \frac{T_P}{T_P + F_P} \tag{2}$$

$$Recall = \frac{T_P}{T_P + F_N} \tag{3}$$

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

The four measurements mentioned above are frequently utilized to assess the effectiveness of HAR. Identification can be described as a true positive (T_P) detection for the category under consideration and a true negative (T_N) detection for all other classes. The misclassification of sensor data can result in false positive or false identification of a category. Specifically, sensor data relating to a particular category could be erroneously identified as relating to another category, leading to false positive (F_P) detection. Conversely, sensor data being connected to a particular category could be mistakenly identified as relating to another category, resulting in false negative (F_N) identification.

IV. Experiments and Findings

This section outlines the conditions of the study. It presents the empirical findings utilized to assess the efficacy of the ResNeXt model for HAR employing IMU sensor data obtained from the LARa dataset.

A. *Experiments Setting*

All experiments conducted in this investigation were executed on the Google Colab Pro infrastructure services utilizing a Tesla V100. The present study employs Python, TensorFlow, Keras, Scikit-Learn, Numpy, and Pandas libraries to carry out the studies.

B. *Experimental Findings*

In order to assess the efficacy of deep learning models that rely on IMU sensors, we carried out a series of investigations to evaluate the ability to recognize the capabilities of both baseline CNN and LSTM models. The ResNeXt model under consideration has undergone hyperparameter tuning through the utilization of the Bayesian optimization technique. This study conducted studies to assess the identification efficacy of deep learning networks using a range of indicators, including accuracy, precision, recall, and F-measure.

Table I presents the F-measure metric obtained from multiple deep learning networks trained on the LARa dataset, specifically for the identification of seven logistics operations.

TABLE I
F-MEASURE OF THE BASELINE DEEP LEARNING MODEL AND THE PROPOSED RESNEXT MODEL IN THIS STUDY

Model	**F-measure (%)**		
	CNN	**LSTM**	**ResNeXt**
Standing	30.3%	40.2%	41.2%
Walking	24.9%	70.4%	72.5%
Moving carts	57.0%	79.9%	79.9%
Handling upwards	20.7%	38.8%	57.2%
Handling centered	74.4%	83.8%	84.1%
Handling downwards	16.5%	71.0%	71.3%
Synchronization	66.7%	79.7%	85.3%

Based on the experimental findings, the ResNeXt model proposed in this study achieved the highest F-measure. Notably, this network demonstrated effective recognition of the "Hand centered" and "Synchronization" actions. However, it showed lower performance in accurately identifying the "Stand" activity, which was the least accurately classified among the seven action categories examined.

V. Conclusion and Future Works

This study aims to explore the effectiveness of using IMU sensors placed on different body parts for categorizing logistical tasks. Two primary deep learning models, CNN and LSTM, were implemented and evaluated. The performance evaluation utilized the LARa dataset, a publicly available benchmark dataset that encompasses a wide range of logistic activities. Additionally, a deep residual neural network called ResNeXt was introduced. The experimental results demonstrate that the ResNeXt model outperforms other baseline deep learning models, as indicated by achieving the highest F-measure for each logistic action.

Future research aims to further develop the ResNeXt architecture by leveraging transfer learning techniques and sensor data from wearable devices for customized identification of human activities.

Acknowledgment

This research project was supported by Thailand Science Research and Innovation Fund; University of Phayao under Grant No. FF66-UoE001; National Science, Research and Innovation Fund (NSRF); and King Mongkut's University of Technology North Bangkok with Contract no. KMUTNB-FF-67-B-09.

References

[1] L. Sanhudo, D. Calvetti, J. P. Martins, N. M. Ramos, P. Mêda, M. C. Gonçalves, and H. Sousa, "Activity classification using accelerometers and machine learning for complex construction worker activities," *Journal of Building Engineering*, vol. 35, p. 102001, 2021.

[2] L. Minh Dang, K. Min, H. Wang, M. Jalil Piran, C. Hee Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognition*, vol. 108, p. 107561, 2020.

[3] F. Niemann, C. Reining, H. Bas, and S. Franke, "Applications of human activity recognition in industrial processes – synergy of human and technology," 2022.

[4] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, ser. IJCAI'16. AAAI Press, 2016, pp. 1533–1540.

[5] S. Saini, A. Juneja, and A. Shrivastava, "Human activity recognition using deep learning: Past, present and future," in *2023 1st International Conference on Intelligent Computing and Research Trends (ICRT)*, 2023, pp. 1–6.

[6] S. Mekruksavanich and A. Jitpattanakul, "Sport-related activity recognition from wearable sensors using bidirectional gru network," *Intelligent Automation & Soft Computing*, vol. 34, no. 3, pp. 1907–1925, 2022.

[7] V. Sharma, B. Sharma, and J. Panda, "Rgb-d dataset: The impact of yoga and gym exercise for human activity recognition," in *2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT)*, 2023, pp. 1–4.

[8] N. Hnoohom, S. Mekruksavanich, and A. Jitpattanakul, "An efficient resnetse architecture for smoking activity recognition from smartwatch," *Intelligent Automation & Soft Computing*, vol. 35, no. 1, pp. 1245–1259, 2023.

[9] M. Fasko, W. Zhao, S. Yang, T. Qiu, and X. Luo, "Towards human activity recognition and objective performance assessment in human patient simulation: A case study," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020, pp. 2702–2707.

[10] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, 2016.

[11] S. Mekruksavanich and A. Jitpattanakul, "Hybrid convolution neural network with channel attention mechanism for sensor-based human activity recognition," *Scientific Reports*, vol. 13, no. 1, p. 12067, 2023.

[12] 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*, 2014, pp. 197–205.

[13] S. Mekruksavanich and A. Jitpattanakul, "A multichannel cnn-lstm network for daily activity recognition using smartwatch sensor data," in *2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering*, 2021, pp. 277–280.

[14] N. Hnoohom, A. Jitpattanakul, I. You, and S. Mekruksavanich, "Deep learning approach for complex activity recognition using heterogeneous sensors from wearable device," in *2021 Research, Invention, and Innovation Congress: Innovation Electricals and Electronics (RI2C)*, 2021, pp. 60–65.

[15] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015.

[16] N. K. Singh and K. S. Suprabhath, "Har using bi-directional lstm with rnn," in *2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)*, 2021, pp. 153–158.

[17] S. Mekruksavanich and A. Jitpattanakul, "Rnn-based deep learning for physical activity recognition using smartwatch sensors: A case study of simple and complex activity recognition," *Mathematical Biosciences and Engineering*, vol. 19, no. 6, pp. 5671–5698, 2022.

[18] S. Park, J. Park, M. Al-masni, M. Al-antari, M. Uddin, and T.-S. Kim, "A depth camera-based human activity recognition via deep learning recurrent neural network for health and social care services," *Procedia Computer Science*, vol. 100, pp. 78–84, 2016.

[19] S. Mekruksavanich and A. Jitpattanakul, "Deep learning approaches for continuous authentication based on activity patterns using mobile sensing," *Sensors*, vol. 21, no. 22, 2021.

[20] M. M. Hossain Shuvo, N. Ahmed, K. Nouduri, and K. Palaniappan, "A hybrid approach for human activity recognition with support vector machine and 1d convolutional neural network," in *2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 2020, pp. 1–5.

[21] S. Mekruksavanich and A. Jitpattanakul, "Smartwatch-based human activity recognition using hybrid lstm network," in *2020 IEEE SENSORS*, 2020, pp. 1–4.

[22] S. Deep and X. Zheng, "Hybrid model featuring cnn and lstm architecture for human activity recognition on smartphone sensor data," in *2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, 2019, pp. 259–264.

[23] N. Hnoohom, A. Jitpattanakul, and S. Mekruksavanich, "Real-life human activity recognition with tri-axial accelerometer data from smartphone using hybrid long short-term memory networks," in *2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, 2020, pp. 1–6.

[24] S. Perez-Gamboa, Q. Sun, and Y. Zhang, "Improved sensor based human activity recognition via hybrid convolutional and recurrent neural networks," in *2021 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL)*, 2021, pp. 1–4.

[25] S. Mekruksavanich, C. Promsakon, and A. Jitpattanakul, "Location-based daily human activity recognition using hybrid deep learning network," in *2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2021, pp. 1–5.

[26] S. Mekruksavanich and A. Jitpattanakul, "Sensor-based complex human activity recognition from smartwatch data using hybrid deep learning network," in *2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, 2021, pp. 1–4.

[27] J. Ward, P. Lukowicz, G. Troster, and T. 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.

[28] J. Zhao and E. Obonyo, "Towards a data-driven approach to injury prevention in construction," in *Advanced Computing Strategies for Engineering*. Cham: Springer International Publishing, 2018, pp. 385–411.

[29] H. Koskimäki, V. Huikari, P. Siirtola, and J. Röning, "Behavior modeling in industrial assembly lines using a wrist-worn inertial measurement unit," *Journal of Ambient Intelligence and Humanized Computing*, vol. 4, pp. 187–194, 2012.

[30] S. Feldhorst, M. Masoudenijad, M. ten Hompel, and G. A. Fink, "Motion classification for analyzing the order picking process using mobile sensors," ser. ICPRAM 2016. Setubal, PRT: SCITEPRESS - Science and Technology Publications, Lda, 2016, pp. 706–713.

[31] 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 *Proceedings of the 5th European Conference on Wireless Sensor Networks*, ser. EWSN'08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 17–33.

[32] J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *Proceedings of the 24th International Conference on Artificial Intelligence*, ser. IJCAI'15. AAAI Press, 2015, pp. 3995–4001.

[33] F. Niemann, C. Reining, F. Moya Rueda, N. R. Nair, J. A. Steffens, G. A. Fink, and M. ten Hompel, "Lara: Creating a dataset for human activity recognition in logistics using semantic attributes," *Sensors*, vol. 20, no. 15, 2020.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.