Alinteri J. of Agr. Sci. (2023) 38(4): 51-58
e-ISSN: 2587-2249
info@alinteridergisi.com

http://dergipark.gov.tr/alinterizbd
http://www.alinteridergisi.com/

DESIGN AND EVALUATION OF PREDICTIVE MODELS FOR HUMAN ACTIVITY RECOGNITION USING DEEP LEARNING

[1] Dr.J.Sravanthi,[2] A.Vennela,[3] B.Rohith,[4] Shaik Rashid,[5] Allamaneni Asmitha
[123]Assistant Professor,[45]Students
Department of CSD
Vaagdevi College of Engineering, Warangal, Telangana

ABSTRACT:

Human Activity Recognition (HAR) has gained significant attention in recent years due to its applications in healthcare, smart homes, fitness tracking, and human-computer interaction. This study focuses on the design and evaluation of predictive models for HAR using deep learning algorithms. By leveraging sensor data from wearable devices, the proposed models aim to classify human activities accurately and efficiently.

Deep learning architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid models are explored to capture spatial and temporal features inherent in HAR datasets. The study utilizes publicly available datasets, such as the UCI HAR and WISDM datasets, for training and testing the models. Comprehensive experiments are conducted to evaluate the performance of these models in terms of accuracy, precision, recall, and computational efficiency.

The results demonstrate that deep learning-based models significantly outperform traditional machine learning approaches by achieving classification accuracies exceeding 95% on benchmark datasets. Moreover, the evaluation highlights the effectiveness of combining CNNs and RNNs for handling the complex temporal dynamics of human activity data.

This research underscores the potential of deep learning in advancing HAR systems and provides insights into the development of scalable, real-time applications for activity monitoring. Future work includes enhancing model generalizability across diverse datasets, addressing class imbalance, and exploring lightweight architectures for deployment on edge devices.

I. INTRODUCTION

Human Activity Recognition (HAR) is a rapidly growing field with applications spanning various domains such as healthcare, fitness tracking, assisted living, and human-computer interaction. The ability to accurately identify and predict human activities from data collected through sensors, such as accelerometers and gyroscopes, has transformative implications for enhancing daily life and optimizing systems in smart environments. Traditional approaches to HAR often relied on handcrafted feature extraction and classical machine learning algorithms, which, while effective, had limitations in capturing the intricate patterns and temporal dependencies inherent in sensor data.

In recent years, deep learning algorithms have revolutionized HAR by automating feature extraction and improving classification accuracy. Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in identifying spatial features, while Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory (LSTM) networks, excel at capturing temporal relationships in sequential data. Additionally, hybrid models combining CNNs and

RNNs have emerged as robust solutions for tackling the complexities of HAR.

This study focuses on designing and evaluating predictive models for HAR using deep learning techniques. By leveraging publicly available datasets, the research aims to develop models capable of accurately recognizing activities such as walking, running, sitting, and lying down. The primary objectives are to analyze the performance of different deep learning architectures, assess their ability to handle real-world scenarios, and propose solutions for practical deployment in applications such as wearable technology and real-time monitoring systems.

The subsequent sections of this paper delve into the methodology, experimental results, and implications of the proposed models. This work aims to contribute to the advancement of HAR systems by providing scalable, accurate, and efficient deep learning-based solutions for activity recognition.

II. BACKGROUND

Human Activity Recognition (HAR) has emerged as a vital research area due to its wide range of applications in healthcare, smart environments, and fitness tracking. Over the years, numerous methods have been developed to improve the accuracy and efficiency of HAR systems. This literature survey provides an overview of traditional approaches, the transition to deep learning methods, and recent advancements in the field.

1. Traditional Methods for HAR

Earlier studies on HAR relied on traditional machine learning algorithms, such as Support Vector Machines (SVM), Decision Trees, and k-Nearest Neighbors (k-NN). These approaches typically required manual feature extraction from sensor data.

Anguita et al. (2013) proposed an SVM-based HAR system using the UCI HAR dataset, achieving high accuracy but with computational complexity due to manual feature engineering.

Kwapisz et al. (2011) utilized the WISDM dataset and applied k-NN and Decision Trees for activity classification, demonstrating the potential of sensor data but facing limitations in generalizability.

While these methods were effective for small datasets and simple activities, they struggled with scalability and recognizing complex activities.

2. Adoption of Deep Learning in HAR

The shift toward deep learning marked a significant improvement in HAR systems by automating feature extraction and handling complex data patterns. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) became widely used for their ability to process spatial and temporal features, respectively.

Ordóñez and Roggen (2016) introduced a deep CNN-LSTM architecture, achieving state-of-the-art results on the Opportunity dataset by combining the strengths of CNNs for spatial feature extraction and LSTMs for capturing temporal dependencies.

Yang et al. (2015) applied deep CNNs to wearable sensor data and demonstrated significant performance improvements over traditional methods, highlighting the scalability of deep learning models.

3. Hybrid Models for Enhanced Performance

Hybrid models combining CNNs and RNNs have been widely explored to leverage both spatial and temporal information.

Murad and Pyun (2017) proposed a hybrid architecture integrating CNNs and Bidirectional LSTMs, which improved HAR accuracy on sequential data.

Hammerla et al. (2016) investigated deep and recurrent networks, emphasizing the importance of data augmentation and domain-specific pretraining for better generalization.

4. Datasets for HAR Research

The availability of public datasets has played a crucial role in advancing HAR research.

The UCI HAR Dataset: Widely used for benchmarking, it includes accelerometer and gyroscope data from smartphones to classify activities such as walking, sitting, and standing.

WISDM Dataset: Focuses on smartphone-based activity recognition, offering a robust dataset for real-time HAR research.

Opportunity Dataset: Captures multimodal sensor data, enabling research on complex activity recognition.

5. Recent Advances in Real-Time HAR

Recent studies have focused on deploying HAR models in real-world scenarios, emphasizing efficiency, robustness, and adaptability.

Chen et al. (2020) developed lightweight CNN architectures for edge devices, enabling real-time HAR on wearable sensors.

Ronao and Cho (2016) explored deep CNNs for smartphone sensor data, achieving competitive results with reduced computational overhead.

6. Challenges in HAR

Despite significant advancements, challenges remain in deploying HAR systems for practical applications. These include handling noisy and imbalanced data, achieving generalizability across diverse datasets, and reducing computational demands for real-time processing.

This survey highlights the evolution of HAR systems from traditional machine learning methods to deep learning-based approaches. By building on this foundation, the current study aims to design and evaluate predictive models that address these challenges, advancing the state-of-the-art in human activity recognition.

III. HUMAN ACTIVITY RECOGNITION USING DEEP LEARNING METHODOLOGIES

This section presents some featured studies that propose models based on CNN, LSTM and hybrid deep learning architectures.

A. CNN

A recent study [24] that utilized a lightweight CNN model for human activity recognition based on wearable devices. It utilized different datasets that obtained data from smartphones of portable sensors. Such datasets include the UCI-HAR, OPPORTUNITY, UNIMIB-SHAR, PAMAP2 and WISDM dataset. Additional studies that utilized data from smartphones along with CNN based human activity recognition models can be found in [25] and [26], where the models performed well with such data.

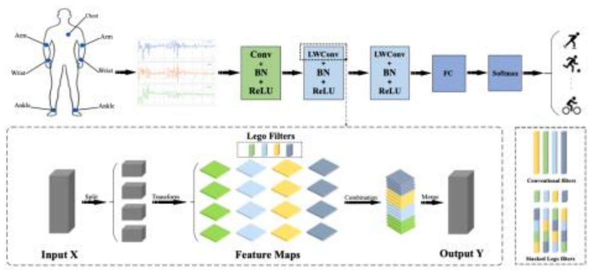

Figure 1. The architecture of a lightweight CNN based model featured in [24]. The model is unique for having a set of lower-dimensional filters which is used as Lego bricks and are stacked for conventional filters resulting in the model being independent of any form of special network structures.

The model is unique for having a set of lower-dimensional filters which is used as Lego bricks and are stacked for conventional filters resulting in the model being independent of any form of special network structures. Signals from the data are split into windows of fixed sizes and overlapping between adjacent windows is tolerated to maintain the continuity of the activities. Ordinary convolution filters that are replaced by the lower dimensional Lego filters makes the model extremely efficient and the Lego filters are simultaneously optimized during the training stage. A straight through estimator (STE) is used as the binary mask for optimizations. To exploit the intermediate feature maps and accelerate convolutions, a classical split transform merge three stage strategy is utilized. ReLU activation functions are used for the different convolution layers and lastly a Softmax function is used to make the activity label prediction based on the network's final output. The local loss functions are implemented using a cross entropy between the prediction of the local liner classifier and its target. The other loss function is based on a kernel of size three by three and a stride and padding of one. The Lego CNN reduces memory and computation costs compared to normal CNN with far lower training parameters and achieved high accuracy through training using the local loss functions. To put the model and datasets to the test, the model was trained on all the listed datasets and evaluated based on overall classification accuracy and the weighted F1 score. As a result, the. Model performed best with the UCI-HAR and WISDM datasets, getting an accuracy of 96.9% and 98.82% and F1 scores of 96.27% and 97.51% respectively.

Another study [27] that proposed a model based on a CNN was one that was focused on unobtrusive activity recognition of elderly people using anonymous binary sensors. A similar study used a CNN based model that is trained using novel spectral data strategies is presented with the goal of detecting freezing of gait in Parkinson's disease patients [28]. The model from [27] was built using a deep convolution network

(DCNN) using an Aruba annotated open dataset that could predict 10 activities that was recorded by a single elderly woman over the span of eight months.

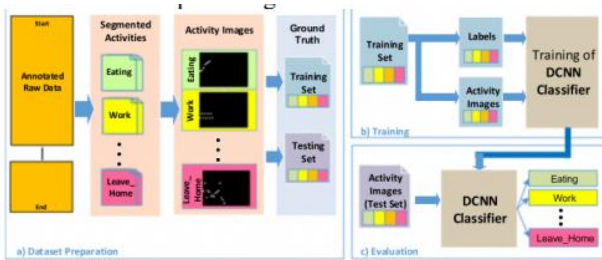

Figure 2. This figure shows the experimental setup [27] used for its DCNN classifier. It focused on using data collected from elderly people using anonymous binary sensors.

As mentioned earlier, it is common for data to be collected using specialized body worn sensors [29], as models proved effective with such techniques in [30]. To preprocess the raw data for the model, the activities are segmented based on their occurrences in the dataset, put through a sliding window to obtain fragmented samples and then lastly the activity was converted into an activity image. The activity image was a simple black and white binary images that had on and off signals that were black and white to represent the activity. The DCNN classifier consisted of three convolution layers, followed by pooling layers that was used for feature extraction. The output of the final max-pooling layer was then flattened and fed to the neurons of the connected layers and in the end the final layer is linked to 10 outputs. The predictive ability of activity was evaluated on accuracy, precision, recall, F1 score and error rate. The activity that had the highest accuracy was going from the bed to the toilet, followed by working, with accuracy scores of 99.99% and 99.85% respectively. On average the DCNN model was able to achieve an accuracy of 98.54% when predicting 10 activities, and 99.23% when predicting 8 activities.

B. LSTM

Bi-directional LSTM (BiLSTM) techniques in human activity recognition have become increasingly popular because of its effectivity with extracting features and making predictions. Models from studies [31] and [32] utilize a BiLSTM for the model's predictive capabilities, while [33] employs it for unique feature extraction techniques. The model utilizes the residual block to extract spatial features from multidimensional signals of MEMS inertial sensors. A CNN based architecture is used in the residual block, as it can extract local spacial features automatically. To get the spatial features from different sensor signals, a 2D CNN residual network with 23 kernels of size 2 x 2 is used. The stride length is 2 and a batch normalization layer is added to speed up training and avoid issues of covariate shift. A ReLU activation function is then used before it goes through another same convolution layer with the same setup except it has a stride of 1. Data was standardized using before being fed into the model. The convolution kernels had dimensions of 2 x 2 because it had the best recognition accuracy. 32 convolution kernels were selected for a balance between the model size and training cost. Cross entropy of the model was minimized using the ADAM optimizer with an optimal learning rate of 0.0003, 0.0006 and 0.00003. A batch side of 64 was used and models were trained a total of 80 times. Forward and backward dependencies of the features are then used by the BiLSTM layer and then features are fed into a Softmax layer for HAR. The model was put to the test one three different datasets: a homemade one, WISDM and PAMAP2. The proposed models also had significantly fewer training parameters and even more ideal results than other deep learning-based models. It performed well for all three datasets, with the homemade dataset getting an accuracy of 96.95%, WISDM getting 97.32% and the PAMAP2 getting 97.15%.

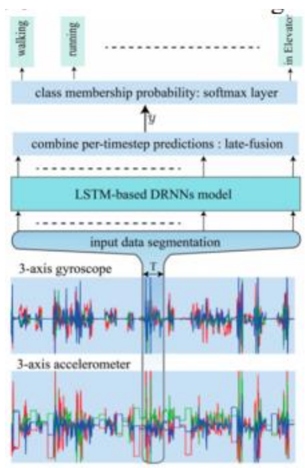

Figure 3. The LSTM based model proposed in [34] that utilizes accelerometer and gyroscope data.

A unidirectional, bidirectional, and cascaded architectures based on LSTM was presented in [34]. The

model uses a deep recurrent neural network for capturing long-range dependencies in variable-length input sequences. This means that it can classify variable-length windows of human activities. 3-axis accelerometer and gyroscope data were used and segmented in time series windows and fed into the model. Upon input, the model outputs a sequence of scores that represents activity labels in which there is a label prediction for the varying time steps. A vector of scores represents the prediction. The SoftMax layer is then used to convert prediction scores into probabilities. Within the LSTM based classifier, three different configurations were built to test three different models. A unidirectional LSTM based DRNN model, a bidirectional LSTM-based DRNN model and lastly a Cascaded Bidirectional and Unidirectional LSTMbased DRNN model. Several datasets were used in testing the effectiveness of the models: the UCI-HAD dataset, the USCHARD dataset, the Opportunity dataset, the Daphnet FOG dataset and lastly the Skoda dataset. Each model was trained using 80% of the data and the remaining 20% was used for validation. The weights of the model were random initially then constantly updated by the mean cross entropy between the ground truth labels and predicted output labels. The Adam optimizer was used for minimizing the cost of backpropagating gradients and updating model parameters. Overall, the unidirectional LSTM based DRNN performed best with the USC-HAD dataset. It had an overall accuracy of 97.8% and an average precision of 97.4%. Furthermore, the hybrid model performed better than traditional machine learning algorithms [36] such as KNN and stand-alone deep learning algorithm-based models such as CNN. Models from [37] and [38] also employ similar techniques of using LSTM layers within the context of an RNN based model for human activity recognition [35].

C. Hybrid Models

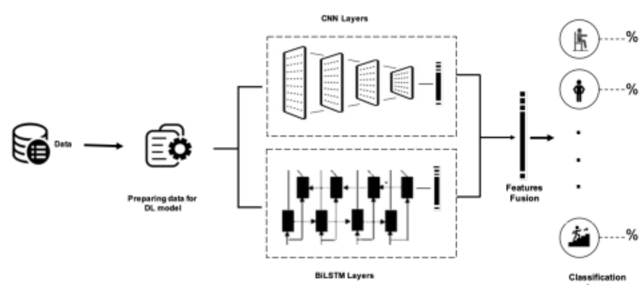

Figure 4. : A hybrid model that utilizes both CNN and (Bi)LSTM layers from [39].

The purpose of another study [39] was to introduce a model for human activity recognition that uses a CNN with varying kernel dimensions that work with a bi-directional long short-term memory (BiLSTM) layer to capture features at different resolutions. This is like what Is seen in proposed models from studies [40], [41] and [42]. These studies all use the combination of both CNN and

LSTM layers in the model for advance feature extraction and processing techniques that make the classification ability of the model more robust. The model from [43] excels at extracting spatial and temporal features from the sensors using the CNN and BiLSTM. The model's input had to be transformed into the out put value using an activation function. The CNN then performed dimensionality reduction of the input data using both maximum and average pooling. The LSTM layer involves automatic learning of high-level features that are related to long term ways across time steps. The BiLSTM part was used for obtaining temporal representation about activity recognition and can access context in forward and backwards directions. The WISDM and UCI datasets were used, and the model was fitted on 30 epochs and had a batch size of 128
samples. The model's accuracy on the WISDM dataset was 98.53% and the UCI dataset was 97.05%.

A total of four models were presented in [45]; the four models include a convolution neural network (CNN) with a Gated Recurrent Unit (GRU), a CNN with a GRU and attention, a CNN with a GRU and a second CNN, and a CNN with Long Short-Term Memory (LSTM) and a second CNN. The study proposes four different human activity recognition models based on deep learning that utilizes channel state information (CSI) in WiFi 902.11n. A public dataset was utilized. Data was collected using the Linux CSI 802.11n tool. This tool is ideal for describing WiFi signals. Actions that occur between a WiFi transmitter and receiver are recorded, and channels display different amplitudes for different actions. The activities that were labeled includes: Lying down, Fall, Walk, Run, Sit down and Stand up. Six people performed each activity 20 times in an indoor office and the data was split into 80% and 90% groups for training and testing. The best performing model was the CNN-GRU model. The CNN-GRU model consists of three parts: input, features extraction and classification. For input, the CSI data is reaped to be suitable inputs to CNNs, so their matrices are reshaped into 1000 x 30 x 3. For feature extraction, two convolutional layers and a GRU layers are used. The input data is put through filters with a size 5 x 5 x 128 kernel and size 1 x 1 stride followed by batch normalization, ReLU activation, average pooling, and dropout with a value of 0.6 is used. Output passes through a flattening layer with time distributed input to convert the data into a vector suitable for the GRU layer. It obtained an accuracy of 99.46%, precision of 99.52% and AUC of 99.90%, outperforming the other three models and other state-of-the-art models. A similar study [46] utilized a deep-neural network-based model that is trained by utilizing transfer learning and shared-weight techniques to classify human activity from cameras. The model which contained Pre-trained CNN and Shared-Weight LSTMRes layers obtained an accuracy of 97.22%.

IV. A SUMMARY AND ANALYSIS OF FEATURED WORKS

Title	*DL based Methodologies*	*Experimental Results*	*Limitations*
Layer-wise Training Convolution al Neural Networks with Smaller Filters for Human Activity Recognition Using Wearable Sensors [24]	A lightweight CNN model for human activity recognition based on wearable devices. The model consists of a set of lower-dimensional filters which is used as Lego bricks and are stacked for conventional filters.	The model performed best with the UCI-HAR and WISDM datasets, getting an accuracy of 96.9% and 98.82% and F1 scores of 96.27% and 97.51%.	The use of smaller Lego filters results in a slight decrease in performance compared to a model based on a traditional CNN.
Unobtrusive Activity Recognition of Elderly People Living Alone Using Anonymous Binary Sensors and DCNN [27]	A model based on a deep convolution network (DCNN) and focused on unobtrusive activity recognition of elderly people using anonymous binary sensors.	On average the DCNN model was able to achieve an accuracy of 98.54% when predicting 10 activities, and 99.23% when predicting 8 activities.	Model's dataset was limited to working with a high-cost setup of binary sensors that recorded the movements for training and validation of the model. Untested with more accessible

Title	DL based Methodologies	Experimental Results	Limitations
			devices such as smartphones or smartwatches.
Human Activity Recognition Based on Residual Network and BiLSTM [33]	Utilized a residual block for extract spatial features from multidimensional signals and bi-directional LSTM (BiLSTM).	It performed well for all three datasets used, with the homemade dataset getting an accuracy of 96.95%, WISDM getting 97.32% and the PAMAP2 getting 97.15%.	Specific activity labels from the datasets that are unbalanced result in lower predicting success. Future work intends to address this to further increase accuracy.
Deep Recurrent Neural Networks for Human Activity Recognition [34]	A unidirectional, bidirectional, and cascaded architectures based on LSTM that uses a deep recurrent neural network for capturing long-range dependencies in variable-length input sequences.	The unidirectional LSTM based DRNN performed best with the USC-HAD dataset. It had an overall accuracy of 97.8% and an average precision of 97.4%.	The model's capabilities were only tested with basic human activities on a small scale and not tested with large scale complex activities.
Sensor-Based Human Activity Recognition with Spatio-Temporal Deep Learning [38]	This model uses a convolution neural network (CNN) with varying kernel dimensions that work with a bi-directional long short-term memory (BiLSTM) layer to capture features at different resolutions.	The model performed well for three datasets, with the homemade dataset getting an accuracy of 96.95%, WISDM getting 97.32% and the PAMAP2 getting 97.15%.	Power and memory usage was not considered for this model, hence the potential for low-powered devices to struggle with such a model.
Utilizing deep learning models in CSI-based human activity recognition [44]	The study proposes four different human activity recognition models based on deep learning that utilizes channel state information (CSI) in WiFi 902.11n.	The best performing model was the CNN-GRU model. It obtained an accuracy of 99.46%, precision of 99.52% and AUC of 99.90%, outperforming the other three models.	No use of denoising in the signals from the data before training.

Table 1 summarizes the features studies that were surveyed as state-of-the-art human activity recognition models that utilize deep learning-based architecture to achieve its predictive capabilities. The papers include a division of models that utilize CNN layers, LSTM layers and hybrid models that utilize more than one algorithm, such as employing both CNN and LSTM layers in the model.

V. CONCLUSION

Human Activity Recognition (HAR) has evolved into a critical research domain with applications in healthcare, fitness tracking, smart environments, and human-computer interaction. This study emphasizes the potential of predictive models based on deep learning algorithms, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid architectures, for accurately classifying human activities using sensor data.

The findings from this study and prior research demonstrate that deep learning models significantly outperform traditional machine learning approaches by automating feature extraction and efficiently capturing spatial and temporal patterns in sensor data. Hybrid models, particularly those combining CNNs and RNNs, further enhance performance by leveraging the strengths of both architectures. Benchmarking on datasets such as UCI HAR and WISDM shows that deep learning-based HAR systems can achieve high accuracy, often exceeding 95%.

However, challenges persist in deploying these models in real-world scenarios, including handling noisy data, ensuring model generalizability across diverse environments, and optimizing computational efficiency for real-time applications. Addressing these issues is essential for enabling the practical implementation of HAR systems in wearable devices, smartphones, and edge computing platforms.

In conclusion, the design and evaluation of predictive models using deep learning provide a promising pathway for advancing HAR technologies. Future research should focus on expanding datasets, addressing data imbalance, and developing lightweight models suitable for real-time and low-power environments. These efforts will contribute to creating scalable, reliable, and practical HAR systems that can benefit individuals and industries worldwide.

REFERENCES

[1] Janelle Mason, Rushit Dave, Prosenjit Chatterjee, Ieschecia GrahamAllen, Albert Esterline, and Kaushik Roy. 2020. An Investigation of Biometric Authentication in the Healthcare Environment. Array 8, (2020), 100042. DOI: https://doi.org/10.1016/j.array.2020.100042

[2] Mousse, Mikaël & Motamed, Cina & Ezin, Eugène. (2017). Percentage of human-occupied areas for fall detection from two views. The Visual Computer. 33. 10.1007/s00371-016-1296-y

[3] Joseph M. Ackerson, Rushit Dave, and Naeem Seliya. 2021. Applications of Recurrent Neural Network for Biometric Authentication & Anomaly Detection. Information 12, 7 (2021), 272. DOI:https://doi.org/10.3390/info12070272

[4] Parker SJ, Strath SJ, Swartz AM. Physical activity measurement in older adults: relationships with mental health. J Aging Phys Act. 2008;16(4):369-380. doi:10.1123/japa.16.4.369

[5] Nyle Siddiqui, Rushit Dave, and Naeem Seliya. 2021. Continuous User Authentication Using Mouse Dynamics, Machine Learning, and Minecraft. 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET) (2021). DOI: https://doi.org/10.48550/arXiv.2110.11080

[6] Sam Strecker, Willem Van Haaften, and Rushit Dave. 2021. An Analysis of IoT Cyber Security Driven by Machine Learning. Algorithms for Intelligent Systems (2021), 725-753. DOI: https://doi.org/10.1007/978-981-16-3246-4_55

[7] Wang, H., Zhang, D., Wang, Y., Ma, J., Wang, Y., & Li, S. (2017). RT-Fall: A Real-Time and Contactless Fall Detection System with Commodity WiFi Devices. IEEE Transactions on Mobile Computing, 16, 511-526.

[8] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook and Z. Yu, "Sensor-Based Activity Recognition," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 6, pp. 790-808, Nov. 2012, doi: 10.1109/TSMCC.2012.2198883.

[9] Liu AA, Xu N, Su YT, Lin H, Hao T, Yang ZX. Single/multi-view human action recognition via regularized multi-task learning. Neurocomputing. 2015; 151:544–553. https://doi.org/10.1016/j.neucom.2014.04. 090

[10] Chaaraoui AA, Climent-Pe´rez P, Flo´rez-Revuelta F. Silhouette-based human action recognition using sequences of key poses. Pattern Recognition Letters. 2013; 34(15):1799–1807. https://doi.org/10.
1016/j.patrec.2013.01.021