Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 204 (2022) 73–80

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Industry Sciences and Computer Science Innovation

Investigation on Human Activity Recognition using Deep Learning

Velliangiri Sarveshwaran[a]*,Iwin Thankumar Joseph[b], Maravarman M[c], Karthikeyan P[d]

[a, c]*B V Raju Institute of Technology,Narasapur, Telangana, india*
[b]*Koneru Lakshmaiah Educational Foundation,Vijayawada, india*
[d]*Jain University, Bnagalore, india*

Abstract

The goal of human activity recognition (HAR) is to describe a people action constructed on a set of sensor readings. Human activity recognition can be classified into two types economic and non-economic. Economic type is used to generate revenue. Non-economic type is used for mental satisfaction. Human activity recognition can be applied in the area of people work evaluations, elderly people care, convalescence, thief detections in a public place, intelligent homes and intelligent traffic. This paper discusses the deep learning model, merits, demerits and dataset used in human activity recognition. Finally, we have summarized essential challenges in HAR using deep supervised and deep unsupervised learning models.

Keywords: Human Activity Recognition; Deep Learning; CNN, LSTM

1. Introduction

Data is one of the most critical aspects of today's scientific world. Data play a significant role in Human Activity Recognition (HAR). This field of study uses various deep learning algorithms to distinguish simple and complex objects dancing, playing, working, and other challenging activities [1]. In the new year's, the field of human movement acknowledgement has developed significantly, mirroring its significance in some high-sway cultural applications, including web-video search and recovery, personal satisfaction gadgets for old individuals, and robot discernment.

* Corresponding author. Tel.: +91-9500519166
E-mail address: velliangiris@gmail.com

10.1016/j.procs.2022.08.009

Convolutional neural network models to learn video depictions, the field is steadily moving towards identifying and anticipating more confusing human exercises affecting numerous individuals, articles, and sub-occasions in different reasonable situations. New significant exploration points and issues are showing up as a result, including (I) dependable spatial-transient limitation of exercises, (ii) start to finish displaying of exercises' sophisticated design and chain of importance, (iii) bunch action acknowledgement, (iv) movement estimating, just as (v) development of massive scope datasets and convolutional models. Human activity recognition is a fascinating problem that can be addressed in ample ways[2][3]. This review is centred around acknowledging actual human exercises dependent on the understanding of sensor information which likewise incorporates one-dimensional time-series information.

Different methods have been implemented during the past years. Various ways are offered to recognize various types of activities, such as whether the individual is running, walking, dancing, jogging, or falling, to list a few. For each technique, multiple methodologies and datasets are employed, with data obtained in various ways such as sensors, accelerometers, gyroscopes, pictures, etc. The desired outcomes by each approach and dataset type are then compared. To categorize, machine learning techniques such as K-nearest neighbours (KNN), support vector machines (SVM), decision trees, hidden Markov models, and unsupervised Deep Learning architectures are majorly used[4][5]. Computers are getting better at solving some very complex problems (like understanding an image) due to the advances in computer vision. Models are being made wherein, if an image is given to the model,it will detect the activity that is present in the model. Deep learning is a subfield of Machine learning. Deep learning can be categorized into two types deep supervised learning and deep unsupervised learning. Human activity recolonization is a complex problem. This problem can be solved using a combination supervised deep learning model and a deep unsupervised learning model. An exciting application of this problem could be identifying objects in videos in real-time[6]. Fig.1 depicts the Overview of Human activity recognition.

The first step in the HAR is data acquisition. Data can be collected from real-time environments like a traffic signal, health care, and sports activity. The collected data can be pre-processed with different pre-processing techniques. The basic pre-processing techniques are the normalization of pixel values and Grayscale conversions. The pre-processing step improves the accuracy of the model. After doing the pre-process, information is passed to the next step (Dividing the data set into two different data sets). The training dataset is passed a deep learning model for training the neural network. Once the model is trained, we can use the testing dataset and test the accuracy.

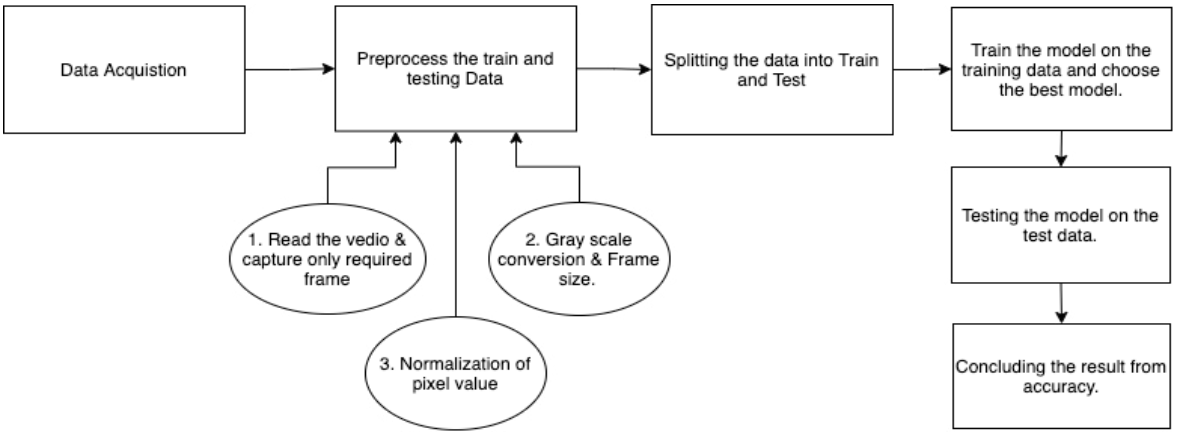

Fig. 1. Overview of Human activity recognition

The HAR takes video as input and outputs the activity performed in that video clip. Generally, HAR can be done in eight different ways i) Single-Frame CNN ii) Late Fusion, iii) Early Fusion iv) Using CNN with LSTM,v) Using Pose Detection and LSTM, vi) Using Optical Flow and CNN's, vii) Using Slow Fast Networks, viii) Using 3D CNN's / Slow Fusion.

Single-Frame CNN: Every single frame is given as input to the CNN model and takes the average of all the discrete probabilities and generate the final probability. This method work well on a small dataset.

Late Fusion: In reality, the Late Fusion method is relatively similar to the Single-Frame CNN method. However, it is a complex than single frame CNN. The sole modification is that in the Single-Frame CNN model, average predicted probabilities occur after the network has completed its job, but in the Late Fusion model, averaging is integrated into deep neural network. As a result, the frames sequence's time-based structure is taken into account. The max pooling, average pooling, or flattening techniques are commonly used to achieve it.

Early Fusion: The video's temporal and channel (RGB) dimensions are fused from the start before giving it to the model.

CNN with LSTM: This technique plans to remove nearby attributes from each casing utilizing convolutional networks. The results of neural organization separate convolutional networks are input into a many-to-one multi-facet LSTM organization, which briefly intertwines the recovered information[7].

Pose Detection and LSTM: Use an off-the-rack act identification model to remove central issues of an individual's body for each edge in the video, then, at that point, send those extricated basic focuses to a LSTM organization to decide the movement being done in the video**.**

Optical Flow and CNN's: Optical flow is a pattern of apparent motion of objects and edges that aids in calculating each pixel's motion vector in a video frame. In motion tracking applications, it works well[8]**.**

SlowFast Networks: This technique, like the last one, employs two parallel streams. In comparison to the other, one stream uses a low-resolution video for a short period. All temporal and spatial operations are carried out in a single network. The top stream, known as the dead branch, uses a low temporal frame rate video and includes many channels at each layer for detailed frame processing. On the other hand, the bottom stream, also known as the short branch, has low channels and works on a high temporal frame rate version of the same movie[9].

3D CNN's / Slow Fusion: This method employs a three-dimensional convolutional network, which allows you to handle both temporal and geographical data. The Slow Fusion Method is another name for this procedure. Contrasting fusion, this approach steadily combines time-based and three-dimensional information at each CNN layer all over the whole network. One disadvantage of this strategy is that increasing the input size dramatically raises the computational and memory requirements [10]**.**

The rest of the paper is organized as follows. We discuss the survey of human activity recognition using the deep learning model in section 2. The open research problem is presented in section 3, and we have concluded the paper in section 4.

2. Related work

This section gives a quick survey of the extant literature on the HAR using a deep learning algorithm. HAR is developed using deep supervised and deep unsupervised models. The classifications are shown in Fig. 2.

Yadav *et al*. developed Changed Inception Time network architecture. The developed model was validated using publicly available datasets such as ARIL, StanWiFi, and SignFi. For WiFi-based activity recognition, the suggested CSITime has obtained an accuracy of 98.20 per cent, 98 percent, and 95.42 per cent on the ARIL, StanWiFi, and SignFi datasets, respectively[11]. Salehzadeh *et al*. presented a Fast Classification of EEG Artifacts (FCEA) deep learning model for classifying EEG artefacts (FCEA) based on a person's physiological activity. The proposed method utilizes the best characteristics of a convolutional neural network and long short-term memory to classify human activity. With the sensory technology frequently utilised in human activity recognition, jaw clenching and head and eye movements actions are difficult to detect. The proposed model work better in terms of F1-score [12].

Agarwal et al. introduce Lightweight Deep Learning Model for human activity recognition that uses less processing power and can be used on edge device Raspberry Pi3. The suggested model's performance is evaluated using data from the participant's six daily activities. The suggested model outperforms numerous existing deep learning techniques[13].

Alazrai et al. designed a complete deep learning system for recognizing human-to-human interactions (HHI) using Wi-Fi signals. This model was evaluated using a well-known CSI dataset collected from 40 distinct pairs of patients while completing 13 humans to human interactions. Across all this, The designed model had a mean accuracy of 86.3 per cent [14].

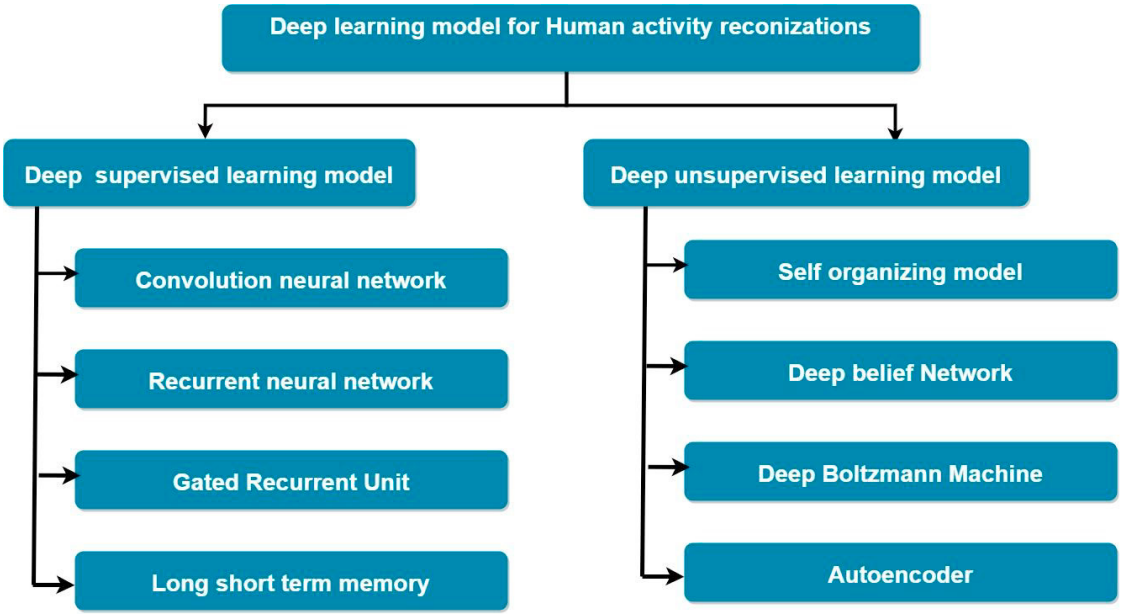

Fig. 2. Different deep learning models for the HAR

Dobhal et al. propose a view-based algorithm for HAR. The proposed deep learning model uses binary Motion images. View-based models work for the 2-D and 3-D data set. Subsampling layers used in this model include the slight invariances to movement, rotations and translations[15].

Hassan et al. discuss the kernel principal component analysis (KPCA) and smartphone sensors to detect human activity. The prosed model first extracts the features, and KPCA and linear discriminant analysis process the extracted features. The proposed model is trained with a Deep Belief Network (DBN) for effective recognition. KPCA provides better accuracy than traditional machine learning models and Artificial Neural networks (ANN)[16]. Janarthanan et al. reported an unsupervised deep learning model for HAR. To reduce the reconstruction error, the proposed model uses the coder architecture. The model is not suitable for the large data set and real-time environment [17]. Jayanthi and Visumathi combined inception ResNet and transfer learning method for HAR using LSTM. Sort the input videos into categories using the LSTM model. The model's accuracy score was linked to the VGG16, ResNet152, and Inception v3 models. The model has trained two different datasets i)UCI 101 and ii) HMDB 51 data sets. ResNet v2 provides the best accuracy score of 92 percent and 91 percent, respectively [18].

Jia et al. introduced multi-frequency and multi-domain HAR based on stepped-frequency continuous-wave radar using deep learning. It uses deep convolutional neural networks (DCNN) and autoencoder for feature extractions. DCNN primary function is to extract micro-Doppler features from a spectrogram. In contrast, auto encodes the main objective is to learn range distribution features. The proposed deep learning model gives 96.42% recognition accuracy [19]. Li et al. present the Robot activity recognition by quality-aware deep reinforcement learning. It uses a policy search model to attain automatic learning of manipulation. The proposed method provides good accuracy for the HAR [20].

Zehra et al. developed HAR through an ensemble of multiple convolutional neural networks. Multiple ensembles and CNN models were trained and tested. The proposed ensemble model gives better accuracy than the traditional models. The main advantage of this method is to extract features that are needed for the model. This model avoids the pre-processing step, so the training and testing time take less time[21]. Ullah et al. discussed the sparse feature learning for HAR. The deep learning model is furnished with inadequate realizing, which assimilates a more prominent number of classes without rolling out a massive improvement in the model's size while supporting the exactness of existing classes.

Moreover, this model is more lightweight than best in class models as it uses FCN-LSTM (Fully convolution organization – Long Short-term Memory). The deep learning Model predicts human exercises like strolling higher up,

strolling the first floor, sitting, standing, and laying (complete six classes). The proposed model is validated using the UCI HAR dataset, giving good accuracy[22].

Subramanian et al. present a method for perceiving the movement of the individual. This movement acknowledgement is one of the critical kinds of wandering wellbeing observing. The patient identified with mental or solid skeletal issues needs encompassing help to screen themselves. This movement acknowledgement model can fill in as an on the web movement observing motor in such use cases. The proposed method provides excellent change location in HAR. There requires an assessment of the progressions between resulting video outlines, not the paired characterization like unaltered. Given the measure of progress, the edge is delegated changed, i.e., the edge addresses another movement [23].

Xiao et al. projected a federated learning model with enhanced feature extraction for HAR. This framework takes perceptive extraction network (PEN) as its component extractor and the FedAvg technique for model load sharing separately. The component and connection organizations successfully investigate neighbourhood highlights and worldwide connections for PEN. Trial results show that PEN acquires the most impressive F1 results among all HAR models. FL includes extractors on four generally perceived datasets [24]. Table. 1 summarizes the merits and demerits of the different deep learning models used in the HAR.

3. Open research problem

In this section, we discuss the open research problem in human activity recognition. Here we discuss the research direction in the four points.

3.1. Video Surveillance for ATMs, Cash Machine Security

ATM security has evolved into a different industry. ATM transactions are rapid and straightforward, but if the machines and the spaces around them are not adequately protected, they might be vulnerable to criminal activities. HAR can be used to classify any unwanted action that happened in the ATM and pass the information's to the police people for this purpose still many researchers are developing a model to solve these issues[25].

3.2. People protection from the thief

The people can be protected from thieves in public places. It is genuinely challenging to watch public places consistently. This way, an automated video observation is required to screen the human exercises progressively and order them as regular and uncommon exercises[26].

3.3. People work evaluations

The people can be recorded in video. We need to develop a deep learning model to detect whether people are working in the workplace or simply chatting with their friends. This system face lot of challenge because of the group of people involved in this problem, and video quality may not be good in the workplace.

Elder people care in healthcare

Remembering headways for the clinical field, innovative progressions have radically worked on our satisfaction, hence pushing the future progressively higher. This has additionally expanded the quantity of the old populace. Like never before, medical care establishments should now focus on an enormous number of old patients, one of the contributing variables in the rising medical services costs. Increasing expenses have incited clinics and other medical care foundations to look for different expense slicing measures to stay cutthroat. There is a need to develop human activity recognition for older people who face many real-time challenges [27].

Table 1 Comparison of Different deep learning models for human activity recognition.

References	Method	Merits	De Merits	Dataset
[11]	Changed Inception Time network architecture	The model works better in terms of accuracy for ARIL, Stan Wi-Fi and SignFi.	The Proposed model may not work for the real-world dataset because the real-world data set has vast interference.	ARIL, StanWiFi, and SignFi
[12]	Fast Classification of EEG Artifacts	Improves classification performance significantly (2–9% for 1-channel data and 8–12% for 2-channel data)	This model was specially developed for the health care sector.	Public Dataset
[14]	The complete deep learning model for Recognizing Human-to-Human Interactions	The developed end to end deep learning model provides 86.3 accuracy for all human-to-human interactions recognition.	The proposed model is not developed for group-to-group interactions. This model will work only for Human-to-Human Interactions.	CSI dataset of HHI
[15]	Binary Motion Image Deep learning	Binary Motion Image Deep learning model gives good accuracy for both 2D and 3D datasets consistent speed of action performed by a human.	The model does not give a reasonable detection rate if more than one person is involved in the 3D image.	MSR action 3D dataset
[16]	Kernel principal component analysis	KPCA outperform Support Vector Machine (SVM) and Artificial Neural Network (ANN)	It provides less accuracy for the real-time data.	Public Dataset
[17]	Unsupervised deep learning assisted reconstructed coder in the on-nodule wearable sensor for HAR	improves the feature selection and extraction using an unsupervised deep learning model	The performances degrade in large datasets with different types of human activities.	WISDM dataset
[18]	Inception ResNet deep transfer learning method for HAR using LSTM	It provides the best accuracy score of 92 per cent and 91 per cent for the different data sets.	It takes a tremendous amount of training time	UCI 101 and HMDB 51 data sets
[19]	Multi-domain HAR based on Stepped-Frequency Continuous-wave radar using deep learning	Developed deep learning model increases the recognition accuracy by 1.3% by additionally introducing the range maps	The proposed model is not developed for group-to-group interactions.	Public Dataset
[21]	HAR using Ensemble Learning of Multiple CNN	It takes less amount of pre-processing time because the proposed model support automatic feature extractions.	Model is not suitable for concurrent activity recognition	WISDM dataset
[22]	Sparse Feature Learning for Human Activity Recognition	It provides long term dependencies	It provides less accuracy for the real-time data.	UCI-HAR dataset

3.4. Human activity recognition for anti-terrorism

Human activity classification and recognition have been a popular study area in anti-terrorism investigations and disaster relief. When electromagnetic waves light a human, his or her moving components produce a Doppler signal. Bodily motions distinguish human micro-Doppler signals, allowing for the classification of human activities[28].

Conclusions

In general, the human activity recognition system is either supervised learning method or unsupervised learning method. An investigative analysis of various deep learning models that were applied in human activity recognition is discussed. Real-world datasets have been taken for investigation. Many open research challenges present that can be a starting point for future research work in human activity recognition. Human activity recognition using deep learning is a real challenge. Still, their undeveloped problem is present. This investigations paper helps the researcher who is interested in starting the research in the field of HAR.

References

[1] H.F. Nweke, Y.W. Teh, M.A. Al-garadi, U.R. Alo, Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges, Expert Syst. Appl. 105 (2018) 233–261. https://doi.org/10.1016/j.eswa.2018.03.056.

[2] T. Zebin, P.J. Scully, K.B. Ozanyan, Human activity recognition with inertial sensors using a deep learning approach, Proc. IEEE Sensors. (2017). https://doi.org/10.1109/ICSENS.2016.7808590.

[3] H. Xu, Z. Huang, J. Wang, Z. Kang, Study on Fast Human Activity Recognition Based on Optimized Feature Selection, Proc. - 2017 16th Int. Symp. Distrib. Comput. Appl. to Business, Eng. Sci. DCABES 2017. 2018-Septe (2017) 109–112. https://doi.org/10.1109/DCABES.2017.31.

[4] P. Hristov, A. Manolova, O. Boumbarov, Deep Learning and SVM-Based Method for Human Activity Recognition with Skeleton Data, 28th Natl. Conf. with Int. Particip. TELECOM 2020 - Proc. (2020) 49–52. https://doi.org/10.1109/TELECOM50385.2020.9299541.

[5] S. Deep, X. Zheng, Leveraging CNN and Transfer Learning for Vision-based Human Activity Recognition, 2019 29th Int. Telecommun. Networks Appl. Conf. ITNAC 2019. (2019) 35–38. https://doi.org/10.1109/ITNAC46935.2019.9078016.

[6] N.H. Friday, M.A. Al-Garadi, G. Mujtaba, U.R. Alo, A. Waqas, Deep learning fusion conceptual frameworks for complex human activity recognition using mobile and wearable sensors, 2018 Int. Conf. Comput. Math. Eng. Technol. Inven. Innov. Integr. Socioecon. Dev. ICoMET 2018 - Proc. 2018-Janua (2018) 1–7. https://doi.org/10.1109/ICOMET.2018.8346364.

[7] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, S.W. Baik, Action Recognition in Video Sequences using Deep Bi-Directional LSTM with CNN Features, IEEE Access. 6 (2017) 1155–1166. https://doi.org/10.1109/ACCESS.2017.2778011.

[8] A.B. Sargano, P. Angelov, Z. Habib, A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition, Appl. Sci. 7 (2017). https://doi.org/10.3390/app7010110.

[9] C. Feichtenhofer, H. Fan, J. Malik, K. He, Slowfast networks for video recognition, Proc. IEEE Int. Conf. Comput. Vis. 2019-Octob (2019) 6201–6210. https://doi.org/10.1109/ICCV.2019.00630.

[10] S. Ji, W. Xu, M. Yang, K. Yu, 3D Convolutional neural networks for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 221–231. https://doi.org/10.1109/TPAMI.2012.59.

[11] S.K. Yadav, S. Sai, A. Gundewar, H. Rathore, K. Tiwari, H.M. Pandey, M. Mathur, CSITime: Privacy-preserving human activity recognition using WiFi channel state information, Neural Networks. 146 (2021) 11–21. https://doi.org/10.1016/j.neunet.2021.11.011.

[12] A. Salehzadeh, A.P. Calitz, J. Greyling, Human activity recognition using deep electroencephalography learning, Biomed. Signal Process. Control. 62 (2020) 102094. https://doi.org/10.1016/j.bspc.2020.102094.

[13] P. Agarwal, M. Alam, A Lightweight Deep Learning Model for Human Activity Recognition on Edge Devices, Procedia Comput. Sci. 167 (2020) 2364–2373. https://doi.org/10.1016/j.procs.2020.03.289.

[14] R. Alazrai, M. Hababeh, B.A. Alsaify, M.Z. Ali, M.I. Daoud, An End-to-End Deep Learning Framework for Recognizing Human-to-Human Interactions Using Wi-Fi Signals, IEEE Access. 8 (2020) 197695–197710. https://doi.org/10.1109/ACCESS.2020.3034849.

[15] T. Dobhal, V. Shitole, G. Thomas, G. Navada, Human Activity Recognition using Binary Motion Image and Deep Learning, Procedia Comput. Sci. 58 (2015) 178–185. https://doi.org/10.1016/j.procs.2015.08.050.

[16] M.M. Hassan, M.Z. Uddin, A. Mohamed, A. Almogren, A robust human activity recognition system using smartphone sensors and deep learning, Futur. Gener. Comput. Syst. 81 (2018) 307–313. https://doi.org/10.1016/j.future.2017.11.029.

[17] R. Janarthanan, S. Doss, S. Baskar, Optimized unsupervised deep learning assisted reconstructed coder in the on-nodule wearable sensor for human activity recognition, Meas. J. Int. Meas. Confed. 164 (2020) 108050. https://doi.org/10.1016/j.measurement.2020.108050.

[18] A. Jeyanthi Suresh, J. Visumathi, Inception ResNet deep transfer learning model for human action recognition using LSTM, Mater. Today Proc. (2020). https://doi.org/10.1016/j.matpr.2020.09.609.

[19] Y. Jia, Y. Guo, G. Wang, R. Song, G. Cui, X. Zhong, Multi-frequency and multi-domain human activity recognition based on SFCW radar using deep learning, Neurocomputing. 444 (2021) 274–287. https://doi.org/10.1016/j.neucom.2020.07.136.

[20] X. Li, J. Zhong, M.M. Kamruzzaman, Complicated robot activity recognition by quality-aware deep reinforcement learning, Futur. Gener. Comput. Syst. 117 (2021) 480–485. https://doi.org/10.1016/j.future.2020.11.017.

[21] N. Zehra, S.H. Azeem, M. Farhan, Human activity recognition through ensemble learning of multiple convolutional neural networks, 2021 55th Annu. Conf. Inf. Sci. Syst. CISS 2021. (2021). https://doi.org/10.1109/CISS50987.2021.9400290.

[22] S. Ullah, D.H. Kim, Sparse feature learning for human activity recognition, Proc. - 2021 IEEE Int. Conf. Big Data Smart Comput. BigComp 2021. (2021) 309–312. https://doi.org/10.1109/BigComp51126.2021.00066.

[23] R.R. Subramanian, V. Vasudevan, A deep genetic algorithm for human activity recognition leveraging fog computing frameworks, J. Vis. Commun. Image Represent. 77 (2021) 103132. https://doi.org/10.1016/j.jvcir.2021.103132.

[24] Z. Xiao, X. Xu, H. Xing, F. Song, X. Wang, B. Zhao, A federated learning system with enhanced feature extraction for human activity recognition, Knowledge-Based Syst. 229 (2021) 107338. https://doi.org/10.1016/j.knosys.2021.107338.

[25] R. Kumar, T. Anand, S. Jalal, Suspicious human activity recognition : a review Suspicious human activity recognition : a review, Artif. Intell. Rev. (2018). https://doi.org/10.1007/s10462-017-9545-7.

[26] M. Paul, S.M.E. Haque, S. Chakraborty, Human detection in surveillance videos and its applications - a review, (2013) 1–16.

[27] P.R. Woznowski, R. King, W. Harwin, I. Craddock, A Human Activity Recognition Framework for Healthcare Applications : Ontology , Labelling Strategies , and Best Practice, (2016) 369–377. https://doi.org/10.5220/0005932503690377.

[28] C. Cheng, F. Ling, S. Guo, G. Cui, Q. Jian, C. Jia, Q. Ran, A Real-time Human Activity Recognition Method for Through-the-Wall Radar, IEEE Natl. Radar Conf. - Proc. 2020-Septe (2020). https://doi.org/10.1109/RadarConf2043947.2020.9266393.