Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 02, February 2024

A Survey on Deep Learning Models for Human Activity Recognition

Archana Vinnod Bansod
Research Scholar, Dept. of CSE
Shri JJTU University,
Rajasthan

Shailesh Kumar
Professor, Dept. of CSE
Gopalan College of Engineering and Management
Bengaluru

***Abstract*— Human activity recognition is the process of predicting an individual's actions based on a sequence of observations of their actions and the environment around them, utilizing various methodologies. It is a dynamic area of research that offers tailored support for a variety of applications and is connected to numerous academic disciplines, including medical services, reliable automation development, and intelligent surveillance systems. An overview of some of the current research techniques for recognizing human activity is given in this publication. It provides an overview of the current state of the art in human activity recognition and presents comparisons of previous studies using a range of features, evaluation criteria, and methods. It also includes the advantages and drawbacks of different techniques to enable researchers to suggest novel strategies.**

***Keywords*— Human Activity Recognition (HAR), Deep Learning, Artificial Intelligence, Computer Vision, Machine Learning**

I. INTRODUCTION

Multiple security cameras are mounted using the Human Activity Recognition (HAR) system to achieve the standard. Identifying one or more environmental factor persons' actions conditionally from a specific range of analyses is the aim of activity recognition. The labelling of movies featuring human activity along with activity categories is part of the Human Activity Recognition process [1]. Activity recognition systems handle a wide range of issues related to video labelling, including things like the type of movies needed, the quality of the exercises completed, and the identification of specific individuals within the activity. While human movement tracking, person detection, and individual identification are sub-levels of the activity recognition system, human activity recognition is utilized to manage regions in conjunction with these functions. The recognition process uses several abstraction levels. In keeping with this, the activity recognition system's point include low-level acts like clearing a table or jumping hurdles, as well as complex movements like jogging or handshaking. Based on the characteristics that are taken from video frames using various algorithms, the human activity is identified. To recognize human activity, these traits are put to the test and learned using machine learning algorithms. Herein lies an explanation of the fundamentals of machine learning techniques, feature extraction, and human activity recognition.

1.1 Overview of Human Activity Recognition:

Due to the limited impact of modern technology, society was astonishingly stress- and competition-free prior to the last three to four decades. Values like contentment and wealth were universal and natural. However, because of rapid progress, shifts in belief systems, and changes in lifestyle, entire scenarios have now transformed. Among the main security concerns facing society and the government are dishonesty, crime, fraud, unethical behavior, etc. Therefore, public spaces monitored by surveillance systems include schools, retail centers, government buildings, theatres, bus and train stations, private offices, homes, etc. Through the use of monitoring cameras, surveillance is an application that offers security to both persons and property. Surveillance refers to the activities of persons being watched in a careful setting. The monitoring apparatus recognizes.

Whether or not the person's actions are morally right. If there is suspicion of activity, it must be stopped or assured action that can control and manage must be taken.

The task of surveillance is repetitive and requires a high level of attentiveness from the observer [2]. Nonetheless, there are a number of scenarios in which inadequate monitoring occurs, failing to record or manage crucial actions. Therefore, the main responsibility of an automatic surveillance system is to recognize and detect people by looking at its metaphors. The most practical and appropriate type of surveillance is video surveillance, which uses monitoring cameras to watch people's movements. When recognition activity runs automatically, the system recognizes people and uses object-based action to steer clear of illicit conduct. An activity that can be carried out following human detection and identification is called an object-based action. For decades, a large number of scientists and researchers have been working in the image processing domain of human recognition.

Another factor contributing to the popularity of video data analysis is the biological visual systems of humans, which are well-suited to process spatiotemporal information. Non-intrusive motion pattern observation is facilitated by video-based monitoring and analysis. When working with human participants, this task needs to be carefully modelled. When there is no possibility for constant observation by visual analysts, video surveillance systems become extremely important. Visual analysts typically work in teams of two to three people to monitor and gather data from several cameras on a continuous basis. When required, they report their findings to the authorities. Therefore, developing a system for

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 02, February 2024

automatically recognizing human actions and creating a more advanced behavioral model for the events taking place in the scene is essential.

Within the field of computer vision, the terms "activity" and "action" are commonly used interchangeably. According to this thesis, "Action" refers to basic movement patterns that a person typically displays for a brief amount of time. Activities include things like walking, running, and waving. On the other hand, "activity" is defined as a complicated series of behaviors in which several people care about one another. Activities are typically defined by far greater temporal lengths. Two people shaking hands, a football team scoring a goal, and multiple people participating in a coordinated bank robbery are a few examples of activities. High-level interactions are required for the recognition of human movements from various angles in a variety of applications, including sports video analysis, motion analysis, human-computer interaction, and video surveillance.

In a video sequence, the typical HAR system can identify, locate, and track moving persons. It can also recognize their actions [3]. The following are the different ways that the Human Activity Recognition system interacts with its surroundings:

- Passive – Passive interaction simply captures and stores the visual information in an organized fashion without performing any analysis.
- Active – Activity interaction controls and adjusts the acquisition device parameters, namely pan, zoom, and tilt effects, depending on the external environment conditions.

  Human Activity Recognition typically has the following three basic steps:
- Detection: Detection involves finding the answer, ‘is there motion present in the scene?’ (Corresponding to a human) and essentially requires low-level processing of images.
- Tracking (feature Extraction): Tracking answers the question: ‘Where is the human moving?’ The tracking is of major importance to Human Activity Recognition. The processing requires to a collect and maintenance of historical data for action recognition and involves the mid- level processing on the history of images. However, sometimes there may be considerable overlap between detection and tracking algorithms.
- Action Recognition / Behavior Understanding (Machine learning Classifier): The process is of high-level vision, which involves interpreting the information derived in the above-mentioned steps to answer the question ‘What is the human doing?’

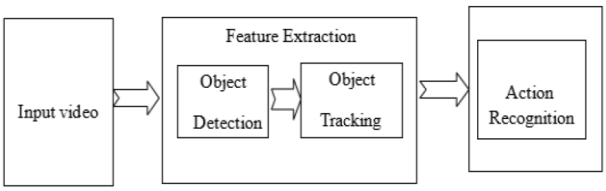

Fig. 1. Architecture of Conventional HAR System

The three stages of a HAR system, which receives films with action as input, are shown in Figure 1. Features are taken out of the input video for the classifier. In the end, the classifier uses the input features to identify the activity.

II. LITERATURE SURVEY

Using a deep learning technique, this section provides a brief overview of the body of literature currently available on the HAR. Both deep supervised and deep unsupervised models are used in the development of HAR. Figure 2 presents the classifications.

A network architecture called Changed Inception Time was created by Yadav et al., [4]. Publicly accessible datasets including ARIL, StanWiFi, and SignFi were used to validate the created model. The proposed CSITime has achieved accuracy of 98.20 percent, 98 percent, and 95.42 percent on the ARIL, StanWiFi, and SignFi datasets, respectively, for WiFi-based activity recognition. A deep learning model called Fast Classification of EEG Artefacts (FCEA) was introduced by Salehzadeh et al. [5] to categorise EEG artefacts (FCEA) according to an individual's physiological activities. The suggested approach classifies human behaviour by combining the best features of long short-term memory and convolutional neural networks. It is challenging to identify activities such as jaw clenching and head and eye movements with the sensory technology commonly used in human activity recognition. F1-score performance is improved by the suggested model.

Agarwal et al. [6] present the Lightweight Deep Learning Model, which can be utilised on the Raspberry Pi 3 edge device and requires less computing resources for human movement identification. The performance of the proposed model is assessed using information from the participant's six daily activities. The proposed model performs better than several deep learning methods already in use. A comprehensive deep learning system was created by Alazrai et al. [7] to identify human-to-human interactions (HHI) through the use of Wi-Fi signals. A well-known CSI dataset that was gathered from 40 different patient couples during 13 human-to-human contacts was used to assess this model. With regard to all of this, the planned model's mean accuracy was 86.3%.

For HAR, Dobhal et al. [8] suggest a view-based method. The binary Motion images are used in the suggested deep learning model. Both the 2-D and 3-D data sets can be used with view-based models. This model makes advantage of subsampling layers for the minor translation, rotation, and movement invariances.

In order to identify human activity, Hassan et al. [9] talk about using smartphone sensors and kernel principal component analysis (KPCA). The proposed model initially extracts the features, which are then processed using linear discriminant analysis and KPCA. For efficient recognition, a Deep Belief Network (DBN) is used to train the suggested model. Compared to artificial neural networks (ANN) and conventional machine learning models, KPCA offers superior accuracy. A deep learning model for HAR that is not supervised was presented by Janarthanan et al. [10], the suggested model makes use of the coder architecture in order to lower the reconstruction error. The big data collection and real-time setting are unsuitable for the model.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 02, February 2024

Jayanthi and Visumathi [11] used LSTM to combine the transfer learning approach for HAR with initial ResNet. The LSTM model is used to classify the input movies. The accuracy score of the model was associated with the Inception v3, ResNet152, and VGG16 models. The UCI 101 and HMDB 51 data sets are the two distinct datasets that the model was trained on. With 92 percent and 91 percent accuracy scores, respectively, ResNet v2 offers the best results.

Deep learning was used by Jia et al. [12] to introduce multi-frequency and multi-domain HAR based on stepped-frequency continuous-wave radar. For feature extraction, autoencoders and deep convolutional neural networks (DCNN) are utilised. Extracting micro-Doppler characteristics from a spectrogram is the main purpose of a DCNN. On the other hand, learning range distribution features is the primary goal of auto encodes. The deep learning model that has been suggested has a 96.42% recognition accuracy.

Robot activity recognition by quality-aware deep reinforcement learning is presented by Li et al. [13], the policy search model is employed to achieve autonomous manipulation learning. For the HAR, the suggested approach offers good accuracy.

HAR was created by Zehra et al. [14] using an ensemble of many convolutional neural networks. We trained and tested a number of ensembles and CNN models. Compared to standard models, the accuracy provided by the suggested ensemble model is superior. This method's primary benefit is its ability to extract features that the model requires. Because this model does not require pre-processing, training and testing take less time. Ullah and colleagues talked about learning sparse features for HAR. The deep learning model's insufficient realizing allows it to incorporate a greater number of classes without significantly increasing the model's size while maintaining the accuracy of the currently available classes. Furthermore, because it employs FCN-LSTM (Fully Convolution Organization – Long Short-term Memory), this model is lighter than best-in-class models. The human exercises of walking upstairs, downstairs, sitting, standing, and lying are predicted by the deep learning model (all six classes). The UCI HAR dataset is used to validate the suggested model, which yields good accuracy [15].

Subramanian et al. offer a technique for recognizing an individual's movement. One of the most important types of wandering wellness watching is this recognition of movement. The patient who has been diagnosed with skeletal or mental health problems requires comprehensive assistance to self-screen. In such application instances, our movement acknowledgement model can serve as a stand-in online movement observation motor. The suggested approach offers HAR's best transition location. Instead of using paired characterization like unaltered, an evaluation of the progressions between the generated video outlines is necessary. The edge is allocated differently based on the progress measure; that is, it takes into account a different movement [16].

A federated learning approach with improved feature extraction was proposed by Xiao et al. for HAR. This framework uses the FedAvg approach for model load sharing and the Perceptive Extraction Network (PEN) as its component extractor. The component and connecting organizations effectively look into local landmarks and global connections for PEN. According to trial results, out of all the HAR models, PEN achieves the most remarkable F1 outcomes. Four widely used datasets have extractors included in FL [17].

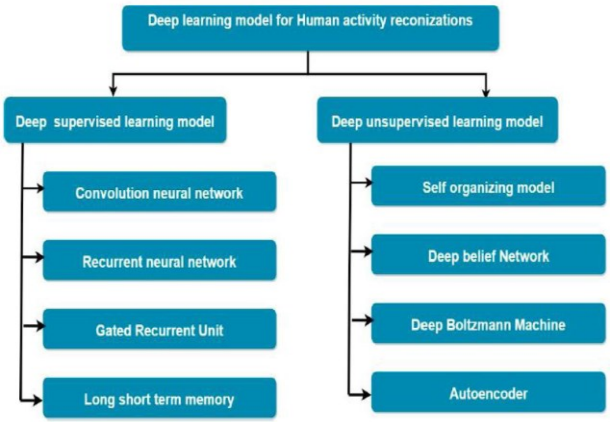

Fig. 2. Different deep learning models for the HAR.

III. GENERALIZED PROPOSED METHODOLOGY FOR HUMAN ACTIVITY RECOGNITION

Identification of human movements in video clips shot in many kinds of environments, such a static background, a background filled with different objects, or a changing and busy background. Research considers a wide range of human activities, from simple ones like sitting, jogging, walking, and standing up to more complicated ones include using a vacuum, relaxing on the couch, or tossing paper.

In the past twenty years, numerous methods have been used in human activity recognition research, employing different approaches and collected data sets. The previous techniques were classified into five groups: appearance-based, vision-based, signal-based, silhouette extraction and modelling, and multi-modal-based. While earlier approaches used manual features extraction techniques, more recent work has focused mostly on deep learning, which makes use of an automatic features extraction process.

The deep neural network-based activity recognition system as a baseline due to its exceptional perform3ance. Specifically, strategies based on convolutional neural networks are provided to improve the overall performance of the HAR system. The proposed techniques are generally applied to various datasets for challenges associated with the identification of human activities. The study work given has been explored with well-known deep neural networks, such as convolutional neural networks and human activity datasets, for comparison with prior work in order to evaluate the current benchmarking of the suggested method with past and future development.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 02, February 2024

IV. CONCLUSION

The human activity recognition system is typically either an unsupervised or supervised learning method. A thorough examination of several deep learning models used for recognising human activities is covered. Investigations have been conducted using real-world datasets. There are now a lot of open research issues that could serve as a springboard for further studies on the recognition of human activities. Deep learning-based human activity recognition is a challenging task. Nevertheless, their unresolved issue still exists. The researcher who wants to begin research in the topic of HAR can benefit from this investigative report.

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] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

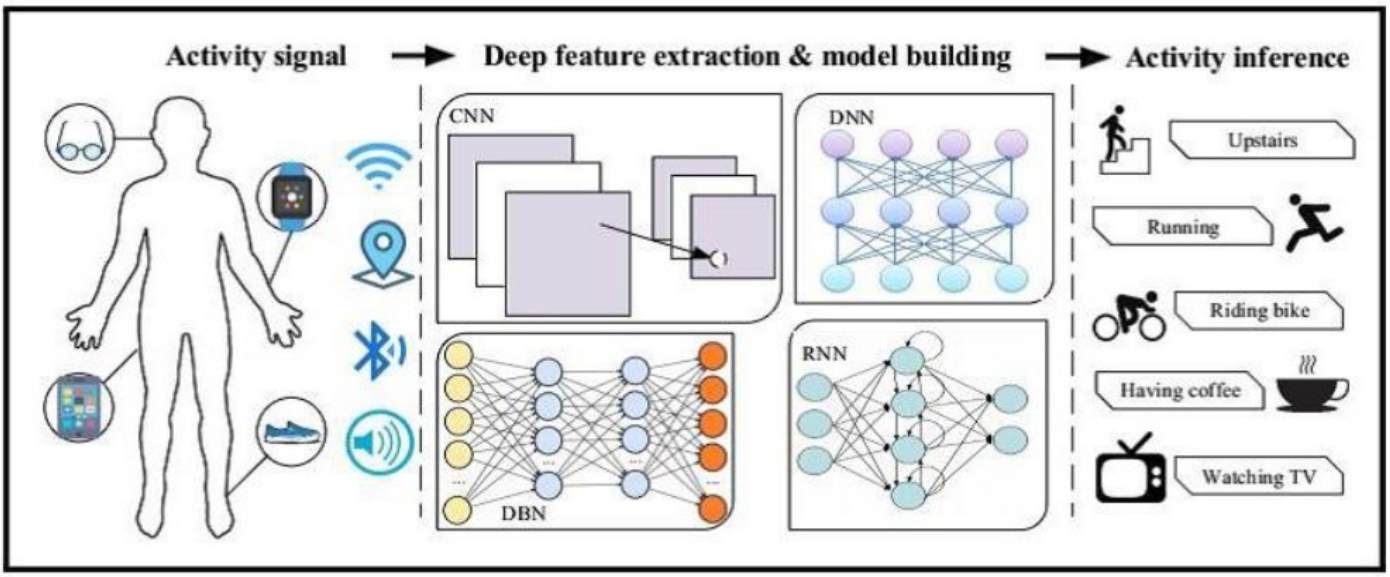

Fig. 3. The whole block diagram of the deep learning-based activity recognition system