Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021)
IEEE Xplore Part Number: CFP21K25-ART

# A Survey on Human Activity Recognition using Sensors and Deep Learning Methods

Khushboo Banjarey
*Information Technology*
*NIT Raipur*
Raipur, India
banjareykhushboo@gmail.com

Dr. Satya Prakash Sahu
*Information Technology*
*NIT Raipur*
Raipur, India
Spsahu.it@nitrr.ac.in

Deepak Kumar Dewangan
*Information Technology*
*NIT Raipur*
Raipur, India
Dkdewangan.phd2018.it@nitrr.ac.in

2021 5th International Conference on Computing Methodologies and Communication (ICCMC) | 978-1-6654-0360-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCMC51019.2021.9418255

***Abstract*— One of the most difficult challenges in the field of computer vision is human activity recognition (HAR). The main purpose of intelligent video system is to determine the actions and activities of the individual. This action monitoring system can be used in a variety of settings, which include human-computer interaction, tracking, security, and health monitoring. Detecting activity in an uncircumscribed territory remains a challenging task with multiple challenges, despite ongoing efforts in this area. Throughout this article, some recent research papers that include different approaches are analyzed to detect different human activities. Wearable devices and phone sensors are required for this task. The vision-based approach has become a prevalent HAR technique, according to the researchers.**

***Keywords—Surveillance, computer-vision, pose-estimation, wearable-devices, smartphone-sensor, HAR.***

## I. INTRODUCTION

The body movements or various limb positions in relation to time and gravity are described as this activity. HAR has become a popular and current research subject for many researchers over the last 20 years, as we all know. However, due to few other unresolved issues including sensor mobility, sensor positioning, contextual setting, and implicit changeability, it remains a difficult task. Human-centric applications can be used to find events in a variety of environments, including home-care assistance, irregular activities, motion detection, physical culture, and health. The HAR framework can identify a significant portion of a person's daily activities as simple or self-created. The majority of HAR structures are based on unsupervised or supervised learning. While unsupervised learning structures have fixed growth norms, supervised learning structures need pre-training using specialized datasets.

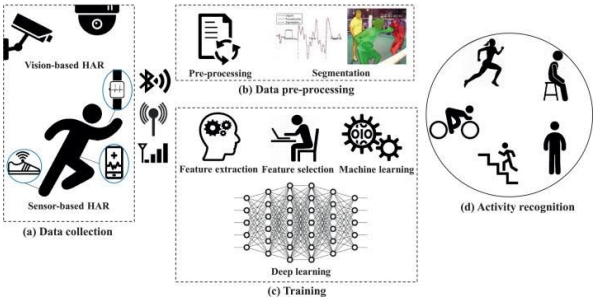


Fig. 1. Basic structure for human activity recognition [1].

The current state of development for various HAR approaches is examined in this paper. It also looks at three technologies: wearing gadget-based approaches, pose-based approaches, and phone sensors. Pose-estimation, which involves the calculation of key-points of the body through neural-network, is used to classify the operations. The pose-based strategy adopts input from sensor data like gyro-scope and accelero-meter, and the phone sensor-based approach takes input from smartphone sensors like gyro-scope and accelero-meter. In the pose-based approach, tasks are classified using pose-estimation, which requires the calculation of key-points of the body through neural-network.

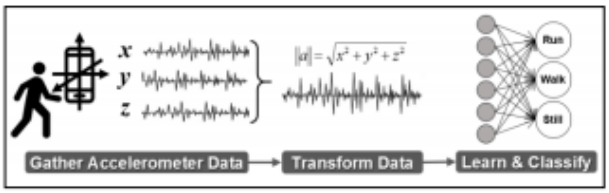


Fig. 2. Flow of CNN-based activity recognition Data [2].

How does HAR work: - First, pre-processing has been done on the given data.

Background removal: It might be essential to separate the person from the background or to exclude any noise.

Bounding box creation: Certain algorithms, especially those in MPPE, produce a bounding box for each person in the image. The human posture is then analyzed separately for each bounding box.

Image registration and camera calibration: - When several cameras are used, image registration is needed. Camera calibration also helps translate recorded ground realities to standard world coordinates in the case of 3D human pose estimation.

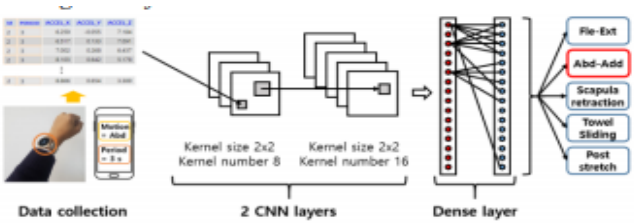


Fig. 3. View of the implemented CNN [3].

HAR typically includes separating joints from a human body and then analyzing a human posture using deep learning algorithms. Key-points are observed from a series of frames, not a single image if the HAR device uses video recordings as a data source. It enables one to gain greater precision because it analyses a person's physical action rather than a static position

## II. RELATED WORKS

Ghazal et al. [4] have been favorable to represented a posture based HAR, basically it utilizes open-present library and forward-feed CNN to anticipate a certainty guide of 18-central issues and employments the dynamic calculation to group the movements of standing and sitting.

Tsai et al. [5] introduced a framework for distinguishing the exercises utilizing a sensor as well as eleven sorts of exercises are perceived via various hidden markov models. The framework is produced for preparing the robot.

Gatt et al. [6] proposed a methodology for recognizing strange conduct, for example, fall identification. The work utilizes Pose Net and Open posture pre-prepared posture assessment model, at that point LSTM and CNN are utilized for movement order.

Bulbul et al.[7] perceive human movement by utilizing diverse grouping and AI approaches, for example,

Packing, k-NN, and so forth for this, they utilized the two cell phone sensors, accelero-meters, and spinner and perceive 6-unique exercises.

Tran et al.[8] introduced a three cell phone sensors-based methodology for the activity-recognition. They utilized SVM for the grouping and ID of movement and streamlined the arrangement framework to recognize the exercises.

RoyChaudhury et al.[9] utilized a solitary cell phone sensor for activity-recognition. They utilize various classifiers to examine the introduced structure. They gone thought about 12-exercises for their completion of work.

## III. METHODOLOGIES AND APPROACHES FOR HAR

The entire process of HAR is can classified in following steps:-

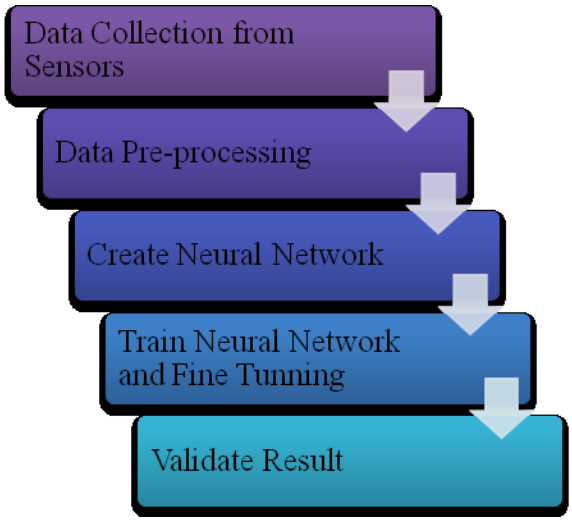


Fig. 4. human activity recognition process.

In order to recognize the activities performed by humans, firstly we need to collect data through sensors or surveillance cameras or any other type of sources, then we have to pre-process the data. In the pre-processing of the data the raw data will be transformed into useful and efficient data. After that we create a neural network model and train the model and the trained neural network will classify the actions.

### *A. Pose Estimation(Vision-Based)*

It is the task of using a machine-learning model to gauge the pose of a person from the given image or video by estimating the endemic area of main joints of body parts. This task is done by determining, placing and tracing of key-points on a given image or video. These key-points signify main articulations like elbow, knee, wrist, waist etc. The main aim of this machine-learning model is to trace these key-points in given picture and recording [10] .

How Pose Estimation Works: - There are two approaches for this task i.e., a bottom-up approach, and a top-down approach.

- In a bottom-up approach the model detects every key-point in a given image and then tries to put-together the groups of key-points into skeletons for different targets.
- And in the top-down approach is the opposite, the network firstly uses an object detector to draw a box around each of the target objects, and then try to guess the key-points inside every cropped area.

However, feature extraction has been utilized in various image processing-based domains and in recent years, computer vision-based image processing [11], [12] and supervised-learning with convolutional-networks (CNNs) has seen huge adoption in intelligent applications [13], [14] and [15].

In recent years, CNN has become most popular deep learning algo. It is essentially used to categorize the pictures and make a group according to their similitude and after that it will do the object recognition.
This is the model that are comprised of two main elements that are convolutional-layer and pooling-layer and the fully-connected-layer in the end also [2].

**Convolutional Layer: -**The main aim of this layer is to bring out the high-level attributes like edges from the given picture. This layer is traditionally the first layer that is chargeable for capturing the low-level attributes such as edges, colors, familiarization etc.
The operations we are doing in the above lines that have two types of result in one in which the resolved characteristics are reduced in comparison of the given input and in the other result the dimensionality will be increases or stay the same.

**Pooling Layer: -**The pooling layer is also chargeable for decreasing the dimensionality of the given features. This is to reduce the computing power that are needed to process the data through dimensionally reduction. It is also useful to remove key features that are rotational and positional unconverted.
There are two types of pooling: - Max-Pooling and Average-Pooling.

- The max pooling returns the maximum value from the portion of the picture that are covered by the kernel. This pooling also clamps the noises. It eliminates the noise movements completely and perform simultaneously with dimensional reduction.
- And the average pooling returns the average value from the portion of the picture that are covered by the kernel. And average pooling only reduces dimensionality as a noise suppressing mechanism.

**Fully Connected Layer: -** It is simply feed forward neural network. The input to this layer is the output from the last pooling-layer or convolutional-layer which is compact and then put in to this layer. The results of convolutional-layer and pooling-layers give to the fully connected neural network architecture that takes the final decision classification.

Some algorithms that are available for pose-estimation are

**Deep Convolutional Neural Network (DCNN): -**

Deep CNN is a type of neural-network with many layers. Deep means that the network has a lot of layers that looks deep stuck of layers in network.

The deep neural network can be used as many fields of supervised machine learning, image classification, text recognition and language translation for convolutional-neural-network.
The architecture of deep CNN contains more than one or many convolutional-layers and one pooling-layer and fully-connected-layer. The working of layers will be same in deep CNN.

We increase the number of layers so we improve the quality of images and rate of convergence. After increasing the number of layers in CNN it is called deep CNN. And using the deep CNN algorithm we can achieved the more accurate result.

**Region-based Convolutional Neural Networks (R-CNN)**

R-CNN's fundamental goal is to take an image as an input and output a series of bounding boxes, each of which includes an entity and the object's type, such as car or pedestrian. R-CNN has also been expanded to execute additional machine vision functions. The following sections detail some of the R-CNN variants that have been developed.
The Region-based Convolutional Network approach (RCNN) combines region proposals with Convolution Neural Networks to create a new method. It uses a drone-mounted camera to monitor objects by detecting text in an image and detecting objects. It achieves high object identification accuracy by classifying object proposals with a deep ConvNet. R-CNN can scale to thousands of object classes without having to rely on estimated techniques like hashing.

**Faster R-CNN:-** Faster R-CNN is a related object tracking algorithm to R-CNN. Compared to R-CNN and Fast R-CNN, this algorithm uses area proposal networks (RPNs), which share the whole image conformable characteristics with the detection network in a cost-effective manner. The Region Proposal Network is an entirely assured network that produces high-quality field resolutions end-to-end while simultaneously predicting the object bound and object score at each position of the object. Is trained, and Fast R-CNN uses it to locate items.

**A recurrent neural network (RNN)**:- It is a type of artificial-neural-network in which nodes are connected in a directed graph that follows a temporal series. This enables it to behave in a temporally complex manner. RNNs, which are generated from feedforward neural networks, can process variable-length sequences of inputs by using their internal state (memory).
Recurrent neural networks (RNNs) get their name from the fact that each component of a chain performs the same operation, with the result dependent on the previous computations.
RNN is recurring, as it executes the same operation in every data input, although the output of the current input is influenced by the previously one. It respects the new input

and the output it has obtained from the previous-input when making a decision.
The way RNNs and feedforward-neural-networks channel knowledge gives them their names.
The knowledge in a feedforward-neural-network only passes in one direction from the input layer to the output layer, passing through the hidden layers. The data flows through the network in a straight line, never moving through same node twice. Feedforward neural networks have no memory of the feedback they receive and are ineffective at anticipating what will happen next. A feedforward network has no concept of time order because it only takes the current input. Except for its training, it has no recollection of what happened in the past.
The information in an RNN passes in a loop. It respects the new feedback as well as what it has heard from previous inputs when making a decision.
However, because of its internal memory, a recurrent neural network may recall certain characters. It creates output, copies it, and then feeds it back into the network.

**Long Short-Term Memory (LSTM): - As** previously said, RNNs cannot memorize data for long periods and continue to neglect their previous inputs. The LSTM is used to solve the problem of vanishing and bursting gradients. They're used to help with short-term memory recall. When new information is applied to RNN, it entirely modifies the previous data. RNN can't say the difference between essential and irrelevant results. When further information is introduced to LSTM, there is just a slight shift in existing information because LSTM includes gates that regulate the flow of information.
The gates assess which data is relevant and will be useful in the future and which data must be discarded. Input, output, and forget gates make up the three gates.
**Forget Gate:** This gate determines which data is relevant and should be saved and which data should be deleted. This gate accepts two inputs: the previous cell's output, and the other is the new cell's entry. The requisite bias and weights are added and multiplied, and the value is then subjected to the sigmoid function. We create a value between 0 and 1 and use it to determine which knowledge to hold. If value is 0, the forget gate will erase the information, so the data must be recalled if the value is 1.
**Input Gate:** This gate adds information to the neuron cell. It determines what values can be applied to a cell using an activation mechanism such as the sigmoid. It generates a list of data that must be entered. Another activation mechanism called tanh is used to do this. It produces values ranging from -1 to 1. The sigmoid feature acts as a buffer, limiting the amount of data that can be applied to a cell.
**Output Gate:** This gate selects critical data from the current cell and shows it as output. It produces a vector of values with a tanh function that varies between -1 and 1. It uses previous output and current input as a controller, which also has a sigmoid function and determines which output values are to be shown.

### *B. Wearable Device Based*

Wearable sensors have gotten extremely famous in numerous applications like clinical, amusement, security, and business fields. They can be very valuable in giving exact and dependable data on individuals' exercises and practices.
Accelerometers are regularly utilized in checking of human action and fundamentally are utilized to gauge quickening along a delicate pivot and over a specific scope of frequencies. They can be utilized for some reasons like discovery of fall development and investigation of body movement. There are a few sorts of accelerometers accessible dependent on piezoelectric, piezoresistive, or variable capacitance techniques for transduction [16].
Wearable gadgets are electronic gadgets that are consolidated into day-by-day use items that are easily worn on a body with added upgrades for putting together large information data consistently. They are furnished with specifically planned movement sensors which can catch the depictions of your everyday action and update that data by adjusting with cell phones or PCs. Sensors can also be arranged around a wearable device to monitor various activities in the vicinity. The majority of sensors can track movement, cerebrum activity, heart activity, and muscle activity. The majority of the time, we see them in health-related gadgets [17].

Most common methods for HAR using wearable device sensors are: -

- Deep Restricted Boltzmann Machine Method.
- Deep Autoencoder.
- Sparse Coding Method.
- Stacked Deep Gaussian Method.

### *C. Smartphone Sensors*

A cell phone sensor is a sort of detecting gadget utilized related to a versatile application on a client's mobile to collect information [18].

Coming up next are a couple of instances of cell phone sensors and their employments:

- An **accelerometer** identifies speeding up, slant and vibration to decide progress and direction.
- A **gyroscope** distinguishes up-down, left-right and orbit around three pivots for more unpredictable direction subtleties.
- A **proximity sensor** identifies when the cell phone is held to the face to settle on or take a

call, so the touch screen can be made inactive to evade unintended action.

- A **finger impression sensor** can empower biometric confirmation for secure gadget, site verification and payment options.

- An **infrared sensor** can be utilized to recognize client progress for motion detection.

The initial step is information gathering that will be finished with the assistance of clients that conveyed a cell phone in their pant pocket while doing easy-going exercises. The subsequent stage will include extraction-Feature Extraction means to diminish the quantity of highlights in a data set by making new highlights from the current ones and afterward disposing of the first highlights. These new decreased arrangement of highlights should then can sum up the vast majority of the data contained in the first arrangement of highlights. [citation from google-towardsdatascience.com]. The last advance of the functional part was exploring different avenues regarding the separated highlights and with three characterization procedures, specifically decision trees, logistic regression, and multi-layer neural network.

## IV. AVAILABLE DATSETS FOR HAR

### *A. UCF-101*

This is one of the most recognized datasets i.e., UCF101 which includes action recognition videos which has been collected from the you-tube. It has 101 categories of real action videos like playing cricket, dancing, running etc. This dataset is extended version of UCF-50 dataset with more videos. Due to large diversity of action videos, this is one of the most challenging data set for action recognition. As we know that various available datasets are not realistic. The ultimate aim of this dataset is to encourage the research work in this field of computer vision. All the videos of this dataset have been grouped into 25 separate groups and each group contains 4 to 7 videos. There is possibility that same grouped video may contains common objects, features and back-ground.

### *B. UCF-50*

This is also one of the datasets of action recognition which includes 50 categories of action videos. This dataset has realistic videos which are collected from you-tube. This dataset i.e., extended version of UCF-11 which has 11 categories of action videos. As we know that various datasets have not realistic action videos. So, the main aim of this dataset is to provide a very challenging, complex and realistic action dataset for research work in computer vision field. Certainly, this dataset is very challenging due to large diversity of real action videos, objects, appearance, motion as well as view-point etc.

### *C. Wisdm*

In this dataset, accelro-meter as well as gyro-scope data has been collected from the recent smart-phones along with the latest smart-watches at the rate of 20Hz. This dataset has been collected from the 51 test-subjects which includes 18 activities of 3 minute duration. There are 4 directories where sensor data collected from the phone as well as watch have stored. For every directory, it has 51 files for 51 corrosponding test-subjects. There is same formate for each and every entry. The des-cription about those attributes are available with the attribute information. There is also example for sensor-data dor 10 second time-window. This dataset is usually used for activity-recognition but it can also be used for behavioural bio-metric architectures because every sensor is collecting information for each and every test-subjects [19].

### *D. UCF Sports Dataset(2007)*

This dataset is specifically created for the various sports activities determination which generally broadcast on tele-visions. This dataset has been collected by one of the University of Florida (computer-vision lab). This dataset has more than 200 videos which are categorized by 9 sports activity such as riding, driving, skating etc. There is one drawback that it covers only 9 sports actions.

### *E. Opportunity*

This Dataset for Human Activity Recognition from Wearable, Object, and Ambient Sensors is a dataset concocted to benchmark human movement acknowledgment calculations like order, programmed information division, sensor fusion, feature extraction, and so on [19].

A subset of this dataset was utilized for the chance Activity Recognition Challenge" coordinated for the 2011 IEEE gathering on Systems, Man and Cybernetics Workshop on "Hearty AI procedures for human movement acknowledgment".

This dataset involves the readings of movement sensors recorded while clients executed run of the mill every day exercises: - Body-worn sensors: 7 inertial estimation units, 12 3-D quickening sensors, 4 3-D confinement data, Item sensors: 12 articles with 3D quickening and 2D pace of turn surrounding sensors: 13 switches and 8 3-D increasing speed sensors, Accounts: 4 clients, 6 runs for each client. Of these, 5 are Activity of Daily Living runs portrayed by a characteristic execution of every day exercises. The sixth

run is a "drill" run, where clients execute a scripted arrangement of exercises.

## V. PREVIOUSLY ACHIEVED ACCURACY

TABLE I. DIFFERENT METHODS FOR HAR AND THEIR ACHIEVED ACCURACY.

| S. No. | *Algorithms* | *Achieved Accuracy* |
|---|---|---|
| 1 | 1D Convolutional Neural Network [2] | 92.71% |
| 2 | Improved dense trajectories (IDT) [20] | 85.9% |
| 3 | IDT with higher dimensional encodings [21] | 87.9% |
| 4 | Slow fusion spatio temporal ConvNet [22] | 65.4% |
| 5 | Implicit CNNs [23] | 89.8% |

The author S. Lee, S. Yoon, H. Cho were used accelerometer data which is gathered through smartphones accelerometer sensor and applied the 1D Convolutional Neural Network method and they have achieved 92.71% accuracy [2].

The author H. Wang and C. Schmid were used the Hollywood2, HMDB51, Olympic Sports and UCF50 datasets and applied the Improved dense trajectories method and they have achieved 85.9% accuracy [20].

The author X. Peng, L. Wang, X. Wang and Y.Qiao were used the HMDB51, UCF50, and UCF101 datasets and applied the IDT with higher dimensional encodings and they have achieved 61.1% accuracy for HMDB51 dataset, 92.3% accuracy for UCF50 dataset, and 87.9% accuracy for UCF101 dataset [21].

The author A. Karpathy, G. Toderici and S. Shetty were used the dataset of 1 million YouTube videos belonging to 487 classes and applied the slow fusion spatio temporal convnet method and achieved 65.4% accuracy [22].

The author Z. Ning, L. Suk-Hwan, L. Eung Joo were used the UCF-101 dataset and applied implicit CNN and they have achieved 89.8% accuracy [23].

## VI. LIMITATIONS

**Model Genre :-** The model genre has a significant influence on AR execution, so it's important to think about what kind of model was used in previous AR studies.

**Number of Subjects & Diversity : -** The number of problems in the dataset has an impact on the quality and durability of the influenced AR-model, as well as the ability to compute the desired results' stability. This is critical because previous findings suggest that the performance of the AR-model of impersonal-models is highly unstable for end-users. To get accurate results, you'll need a large enough number of subjects to calculate.

**Methodology of Data Collection: -** The method of data collection is critical and must be noted. There are three types of AR collections at higher levels, despite the fact that various odds can be created.

- **Completely natural: -** Subjects describe their daily routines without altering their behaviour.

- **Semi-natural: -** Subjects work in their areas but make minor changes to their behaviour, such as ensuring that they walk or engage in other particular activities.

- **Laboratory:-** In a laboratory setting, subjects participate in organised activities. The activity should include a documentation of the data format's methodology as well as an attempt to map the sampling technique to one of these domains. It's crucial to document this because productivity will be better in a lab environment, so this knowledge should be made public for fairness' sake. Some studies include this data, while others do not. We've launched two datasets: one for "Activity Recognition," which uses moderately AR data collection, and another for "Actitracker," which uses only natural data collection through our Actitracker app.

**Sensors : -** It is necessary to characterise all aspects of the sensor that affect AR data. Some of these are often properly identified, such as the type of sensor (accelerometer, gyroscope, GPS), and the number of sensors. The precise location of each sensor, on the other hand, is also not specified, and the orient is often not specified. For instance, if a smartphone is carried in a pants pocket, which pocket is it carried in? What is the orientation of the phone? While this knowledge may not appear to be important, our experience has shown that these factors have an impact on AR performance. As a result, we often specify the precise location and orientation in our technical parameters; however, we, as well as many others, failed to include these details in the methods section of our research articles. Individual models may not care about these information because they are generated — but even in these situations, the model may have issues if the position and orientation change over time. If protocols obstruct location and orientation but aren't recorded, others won't be able to replicate the task and draw the correct conclusions.

## VII. CONCLUSION AND FUTURE SCOPE

A study has been finished on some chosen recent research papers on different Human Activity Recognition innovations. We have ordered these advances into three principal classes, to be specific HAR utilizing 1. Present assessment (vision-based), 2. Cell phone sensors, and 3. Wearable sensors. However, a few calculations have low activity acknowledgment rates. Further, examination is required in this field to procure the improved exactness and build the quantity of exercises recognized. Furthermore, it is observed that the vision-based approach is a very common method for human activity recognition.

## REFERENCES

[1] L. Minh Dang, Kyungbok Min, Hanxiang Wang, Md. Jalil Piran, Cheol Hee Lee, Hyeonjoon Moon, Sensor-based and vision-based human activity recognition:A comprehensive survey,Pattern Recognition,Volume 108,2020,107561,ISSN 0031-3203,https://doi.org/10.1016/j.patcog.2020.107561.

[2] S. M. Lee, S. M. Yoon, and H. Cho, "Human activity recognition from accelerometer data using Convolutional Neural Network," *2017 IEEE Int. Conf. Big Data Smart Comput. BigComp 2017*, pp. 131–134, 2017, doi: 10.1109/BIGCOMP.2017.7881728.

[3] K. S. Lee, S. Chae, and H. S. Park, "Optimal time-window derivation for human-activity recognition based on convolutional neural networks of repeated rehabilitation motions," *IEEE Int. Conf. Rehabil. Robot.*, vol. 2019-June, pp. 583–586, 2019, doi: 10.1109/ICORR.2019.8779475.

[4] S. Ghazal and U. S. Khan, "Human posture classification using skeleton information," *2018 Int. Conf. Comput. Math. Eng. Technol. Inven. Innov. Integr. Socioecon. Dev. iCoMET 2018 - Proc.*, vol. 2018-Janua, pp. 1–4, 2018, doi: 10.1109/ICOMET.2018.8346407.

[5] A. C. Tsai, Y. Y. Ou, C. A. Sun, and J. F. Wang, "VQ-HMM classifier for human activity recognition based on R-GBD sensor," *Proc. 2017 Int. Conf. Orange Technol. ICOT 2017*, vol. 2018-Janua, pp. 201–204, 2018, doi: 10.1109/ICOT.2017.8336122.

[6] T. Gatt, D. Seychell, and A. Dingli, "Detecting human abnormal behaviour through a video generated model," *Int. Symp. Image Signal Process. Anal. ISPA*, vol. 2019-Septe, pp. 264–270, 2019, doi: 10.1109/ISPA.2019.8868795.

[7] E. Bulbul, A. Cetin, and I. A. Dogru, "Human Activity Recognition Using Smartphones," *ISMSIT 2018 - 2nd Int. Symp. Multidiscip. Stud. Innov. Technol. Proc.*, pp. 2–7, 2018, doi: 10.1109/ISMSIT.2018.8567275.

[8] D. N. Tran and D. D. Phan, "Human Activities Recognition in Android Smartphone Using Support Vector Machine," *Proc. - Int. Conf. Intell. Syst. Model. Simulation, ISMS*, vol. 0, pp. 64–68, 2016, doi: 10.1109/ISMS.2016.51.

[9] I. Roychowdhury, J. Saha, and C. Chowdhury, "Detailed Activity Recognition with Smartphones," *Proc. 5th Int. Conf. Emerg. Appl. Inf. Technol. EAIT 2018*, 2018, doi: 10.1109/EAIT.2018.8470425.

[10] A. Gupta, K. Gupta, K. Gupta, and K. Gupta, "A Survey on Human Activity Recognition and Classification," *Proc. 2020 IEEE Int. Conf. Commun. Signal Process. ICCSP 2020*, pp. 915–919, 2020, doi: 10.1109/ICCSP48568.2020.9182416.

[11] D. K. Dewangan and S. P. Sahu, "Driving Behaviour Analysis of Intelligent Vehicle System for Lane Detection using Vision-Sensor," *IEEE Sens. J.*, vol. 21, no. 5, pp. 6367–6375, 2020, doi: 10.1109/JSEN.2020.3037340.

[12] D. K. Dewangan and S. P. Sahu, "Real time object tracking for intelligent vehicle," *2020 1st Int. Conf. Power, Control Comput. Technol. ICPC2T 2020*, pp. 134–138, 2020, doi: 10.1109/ICPC2T48082.2020.9071478.

[13] D. K. Dewangan and S. P. Sahu, "Deep Learning-Based Speed Bump Detection Model for Intelligent Vehicle System Using Raspberry Pi," *IEEE Sens. J.*, vol. 21, no. 3, pp. 3570–3578, 2021, doi: 10.1109/JSEN.2020.3027097.

[14] Dewangan, D.K. and Sahu, S.P. (2021), PotNet: Pothole detection for autonomous vehicle system using convolutional neural network. Electron. Lett.. https://doi.org/10.1049/ell2.12062.

[15] D. K. Dewangan and S. P. Sahu, "RCNet: road classification convolutional neural networks for intelligent vehicle system," *Intell. Serv. Robot.*, no. 0123456789, 2021, doi: 10.1007/s11370-020-00343-6.

[16] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sens. J.*, vol. 15, no. 3, pp. 1321–1330, 2015, doi: 10.1109/JSEN.2014.2370945.

[17] Ó. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surv. Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013, doi: 10.1109/SURV.2012.110112.00192.

[18] C. Y. Shan, P. Y. Han, and O. S. Yin, "Deep Analysis for Smartphone-based Human Activity Recognition," *2020 8th Int. Conf. Inf. Commun. Technol. ICoICT 2020*, 2020, doi: 10.1109/ICoICT49345.2020.9166229.

[19] K. Xia, J. Huang, and H. Wang, "LSTM-CNN Architecture for Human Activity Recognition," *IEEE Access*, vol. 8, pp. 56855–56866, 2020, doi: 10.1109/ACCESS.2020.2982225.

[20] H. Wang and C. Schmid, "Action recognition with improved trajectories," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 3551–3558, 2013, doi: 10.1109/ICCV.2013.441.

[21] X. Peng, L. Wang, X. Wang, and Y. Qiao, "Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice," *Comput. Vis. Image Underst.*, vol. 150, pp. 109–125, 2016, doi: 10.1016/j.cviu.2016.03.013.

[22] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and F. F. Li, "Large-scale video classification with convolutional neural networks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1725–1732, 2014, doi: 10.1109/CVPR.2014.223.

[23] Z. Ning, L. Suk-Hwan, and L. Eung-Joo, "Human Activity Recognition Based on Loss-Net Fusion Domain Convolutional Neural Networks," *2019 IEEE Int. Conf. Comput. Commun. Eng. ICCCE 2019*, pp. 146–149, 2019, doi: 10.1109/ICCCE48422.2019.9010800.