# HUMAN ACTIVITY RECOGNITION USING SMARTPHONE SENSORS

MAJOR PROJECT REPORT

***Submitted in partial fulfilment of the requirements for the award of the degree***

***of***

**BACHELOR OF TECHNOLOGY**

***in***

**COMPUTER SCIENCE AND ENGINEERING**

by


| Divye Gupta | Aarnav JP | Muskaan Bhardwaj | Mayank Hora |
|---|---|---|---|
| 02811502720 | 00211502720 | 06011502720 | 05611502720 |

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**BHARATI VIDYAPEETH'S COLLEGE OF ENGINEERING**

**(AFFILIATED TO GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY, DELHI) DELHI – 110063**


**May 2024**

# CANDIDATE'S DECLARATION

It is hereby certified that the work which is being presented in the B. Tech Major Project report entitled **"HUMAN ACTIVITY RECOGNITION USING SMARTPHONE SENSORS"** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** and submitted in the **Department of Computer Science and Engineering** of **BHARATI VIDYAPEETH'S COLLEGE OF ENGINEERING, New Delhi (Affiliated to Guru Gobind Singh Indraprastha University, Delhi)** is an authentic record of our own work carried out during a period from **February 2024 to May 2024** under the guidance of **Ms. Rachna Narula, Assistant Professor.**

The matter presented in the B. Tech Major Project Report has not been submitted by me for the award of any other degree at this or any other Institute.

| Divye Gupta | Aarnav JP | Muskaan Bhardwaj | Mayank Hora |
|---|---|---|---|
| 02811502720 | 00211502720 | 06011502720 | 05611502720 |

This is to certify that the above statement made by the candidates is correct to the best of my knowledge. He/She/They are permitted to appear in the External Major Project Examination

| | |
|---|---|
| **Ms. Rachna Narula** | **Dr. Deepika Kumar** |
| **(Assistant Professor)** | **(Head of Department,CSE)** |

# ABSTRACT

Human activity recognition (HAR) has become increasingly important with the widespread use of smartphones and their sensors, coupled with advancements in AI technology. This report aims to analyze the three pillars of HAR: smartphone sensors, AI technology, and their applications from 2011 to 2021. The study also provides recommendations for improved HAR design, reliability, and stability. The five key findings of the analysis are as follows: (1) HAR relies on smartphones as acquisition devices, utilizing sensors such as accelerometers, gyroscopes, magnetometers, GPS, and proximity sensors. (2) HAR has made significant strides in the healthcare industry, enabling remote patient monitoring, sleep tracking, and overall well-being assessment. (3) Hybrid AI models in HAR are in their early stages and require further development to establish stable and reliable designs. (4) Limited attention has been given to detecting irregularities or anomalies during human actions, which could be crucial for applications like fall detection and injury prevention. (5) Predictive modeling for future activity recognition based on smartphone sensor data remains largely unexplored. The HAR industry will continue to improve by leveraging advancements in smartphone sensor technology, AI algorithms, and innovative applications. AI will serve as a driving force, empowering more accurate and reliable HAR systems. The proposed review sheds light on the potential for enhancing HAR systems through comprehensive analysis of the three pillars and provides insights for future research and development in the field.

# ACKNOWLEDGEMENT

We express our deep gratitude to **Ms. Rachna Narula**, Assistant Professor, Department of Computer Science and Engineering for her valuable guidance and suggestion throughout our project work.

We would like to extend our sincere thanks to the **Head of the Department, Dr. Deepika Kumar** for his time-to-time suggestions to complete our project work. We are also thankful to **Prof. (Dr.) Dharmender Saini, Principal** for providing us with the facilities to carry out our project work.

| | | | |
|---|---|---|---|
| Divye Gupta | Aarnav JP | Muskaan Bhardwaj | Mayank Hora |
| 02811502720 | 00211502720 | 06011502720 | 05611502720 |

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| *HAR*: | Human Activity Recognition |
| *CNN*: | Convolutional Neural Network |
| *LSTM*: | Long Short Term Memory |
| *AI*: | Artificial Intelligence |
| *DL*: | Deep Learning |

# CHAPTER 1: INTRODUCTION

## 1.1 INTRODUCTION

Over the recent years, Human Activity Recognition (HAR) has become a topic of significant importance and research. HAR simply refers to the task of measuring a person’s physical activity with the help of objective technology. This task is extremely challenging owing to the complexity and diversity of human activities, which can be simplified by sub-categorizing. HAR has become an interesting research field over the course of time, especially because of the spread of electronic devices like mobile phones [8,15,26], smart cell phones, and video cameras [30] in our daily lives. Adding further, the progress of deep learning and other algorithms has made it possible for researchers to use HAR in many fields including sports, health, well-being, etc. For example, HAR is one of the most promising resources for helping older people [10, 12] support their cognitive and physical function through day-to-day activities.

With the current developments in some areas using deep learning, a field that has not received much attention is HAR. In the normal case for HAR, the user has a device (it can be a standalone sensor, smartwatch, smartphone, etc.) i.e., fitted with a gyroscope and accelerometer sensor [2,8,10], always sends sensor data to a server that listens to that to enable continuous activity monitoring. Changes to these structures exist, especially with modern smart devices [16,17,29,35] having the skills to do activity recognition as well as self-monitoring [14]. These have better processing units, more significant memories, and better senses.

HAR programs can be divided into two main groups, namely video-based systems, and sensor-based systems [34]. In video-based HAR systems, cameras are used to record images or videos to monitor human behavior, while sensors in the body and surrounding areas are used by HAR-based sensory systems to capture and record human activity data [7,8,32]. Due to the privacy issue raised by the installation of cameras in the participants' surroundings, requests to monitor daily activities are governed by sensor-based systems [18,35]. In addition, another benefit of nerves is their proliferation.

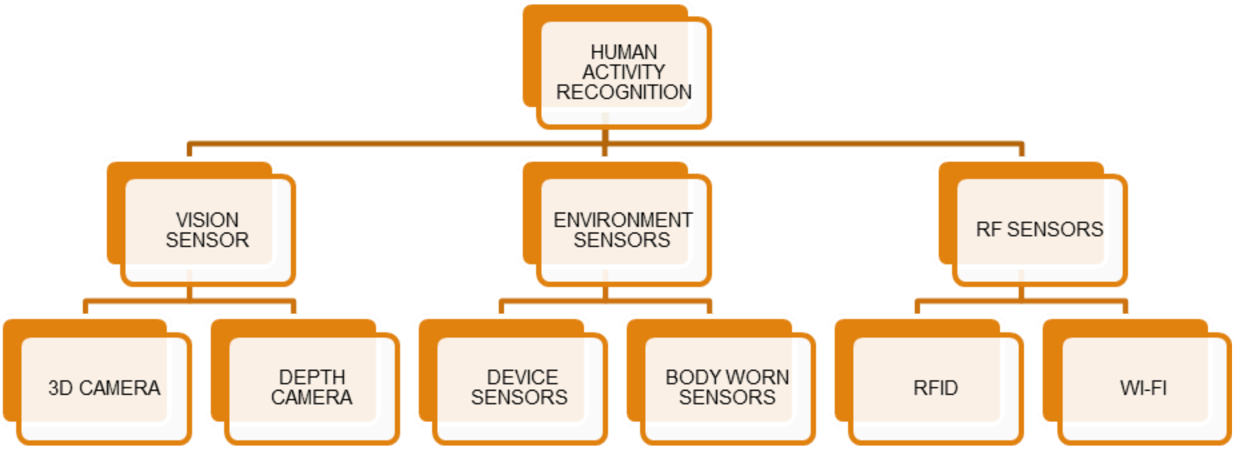


**Fig. 1.1 Classification of Human Activity Recognition**

As a result of the extended existence of smart devices with various sensors, it is possible to insert sensors into portable devices [2,8,11,17,29,36,40], such as mirrors, phones, and watches, as well as intangibles, such as cars, furniture, and walls.

The objective of HAR is to consider human activity in regulated and unregulated environments. Even though there are numerous applications of HAR, there are various challenges faced, like:

(1) complexity and variety of day-to-day activities
(2) intra and inter-subject variability for the same operations
(3) trade-off between privacy and performance
(4) embedded and portable system calculation reliability

For processing and analyzing of HAR, we obtain the data from two key sources: training and testing datasets, ambient sensors, and embedded sensors.

Deep learning is a highly preferred technique these days [3,13]. Whether it comes to natural language processing or computational vision, deep learning has increased prominence in modeling high-level abstractions of nuanced data.

The advantages of deep learning in human activity recognition are

- It is one of the most appealing attributes Layer by layer deep model architectures make it possible to learn from easy-to-abstract functionality. Several advanced computing tools like Graphics Processing Units (GPUs) often allow deep models to learn descriptive functions from complex data.
- Various neural network architectures represent multifaceted functions. For instance, Convolutional neural networks (CNNs) can capture multimodal sensory input locally and the local translation invariance is accurate.
- Deep neural networking can be detachable and scalable into interconnected networks with a global optimization feature that promotes various deep learning strategies learning.

As per the research of [27], a mobile application named FonLog can be used as a tool for data collection to detect human activities in nursing homes and other Healthcare areas. The app consists of various features like recording activity targets, recognition feedback, detailed records, a user-friendly interface, and functionality of offline operation that can help collect the data of various activities effectively.

The model proposed by [35] uses the approach of a single triaxial accelerometer which doesn’t impose any strict restrictions on the smartphone’s location though it does take into consideration some typical locations where the users carry their smartphones, and this activity recognition system does not demand a firm connection of the smartphone with the subject. Yet it serves as a relevant example to exhibit the role, challenges, and a few main potential effects that these smartphones have now and will have in the future in mHealth.

The works of [6] tend to research the significance of accelerometer and gyroscope sensors and their conjunction for automatic human activity recognition using artificial neural networks.

According to the experiment conducted, the discussed sensors may be used to identify human actions individually. On combining both the sensors together, they performed better than when they were used individually, however merging many sensors together may lead to significant difficulties due to the restrictions on the smartphone's battery. Action recognition requires continuous signals from the smartphone.

The study by [27] proposes a mobile application named FonLog which can be used as a data collection tool in action recognition of humans for nursing services. This app has the necessary characteristics for efficient data collection which is gathered by the feedback from nursing staff which includes recording action targets, instant activity, recognition feedback, detailed records that are customizable, and also offline functionality.

The model proposed by [4] introduces a technique that focuses on enhancing the performance of the tracking application used for the COVID-19 virus with the help of human activity recognition (HAR) and a classifier based on Convolutional neural networks is provided for the same.

In the proposed method the raw data from the accelerometer signals of a smartphone is organized to form an image known as HAR-image which is further used as fingerprints of the in-progress activity. Experiments dealing with analyzing real data have shown that HAR-Images can be used as beneficial characteristics for human activity recognition.

An IoT technology-based intelligent m-Healthcare system devised by [37] that used data mining techniques to provide extensive human activity recognition has been proposed in this research. A reliable and accurate IoT technology-based human action recognition model is developed with a client-dependent data mining strategy for offline human action recognition.

Another domain that is proving to emerge manifolds in the industry of information technology is the Internet of Things (IoT). IoT can be applied effectively in the modern health care field, by actively monitoring the daily activities of patients and elderly people [7, 8, 9]. This enables the role of HAR with IoT significant for smart healthcare devices. One major asset provided by

IoT to the Healthcare monitoring system is the technology of wearable sensors. In addition, the incorporation of IoT with health care has led to the development of smart applications like m-Healthcare and other monitoring systems [10].

With the current developments in some areas using deep learning, a field that has not received much attention is HAR.

In the normal case for HAR, the user has a device (it can be a standalone sensor, smartwatch, smartphone, etc.) i.e. fitted with a gyroscope and accelerometer sensor, always sends sensor data to a server that listens to that to enable continuous activity monitoring [41]. Changes to these structures exist, especially with modern smart devices having the skills to do activity recognition as well as self-monitoring. These have better processing units, bigger memories, and better senses.

In the case of nursing activities, observing the pattern and paying attention to methods are some of the ways that are beneficial in increasing the effectiveness of health care [27, 19] One such instance can be including only the essential activities in the exercise routine of the patient in order to aid earlier discharge and avoid excessive and unnecessary work. Various researchers have adopted such practices with the use of mobile sensors like accelerometers, gyroscopes, etc. in related fields of Healthcare and nursing activities.

As per research by [27], it was found that several key features are required for effective activity recognition such as a user-friendly interface of an application, recording target persons, interface for recognition and feedback, detailed records, and the functionality of offline database and operation.

HAR programs can be divided into two main groups, namely video-based systems and sensor-based systems. In video-based HAR systems, cameras are used to record images or videos to monitor human behavior, while sensors in the body and surrounding areas are used by HAR-based sensory systems to capture and record human activity data. Due to the privacy issue raised by the installation of cameras in the participants' surroundings, requests to monitor daily activities are governed by sensor-based systems. In addition, another benefit of nerves is their proliferation. As a result of the extended existence of smart devices with various sensors, it is possible to insert sensors into portable devices, such as mirrors, phones, and watches, as well as intangibles, such as cars, furniture, and walls. Nowadays, sensors are widely installed in our environment, recording information about human activity, and do not invade the privacy of the user by any means.

## 1.2 OBJECTIVE

The objective of HAR is to consider the human activity as a regulated and unregulated environment. It helps in classifying a user's movement by studying a series of measurements captured by sensors or other sources.
With the advancements in technology, there are numerous ways to help monitor the movements of a person. It can be a smartwatch [8,15], a pulsometer, a smartphone, etc.

Typically, this is done by following the sliding window method of the fixed length to remove the features. Here 2 parameters must be fixed: window size and shift HAR have a lot of benefits [13] attached to it in the field of health, wellness, or sports

**1. Health Monitoring -**
Analyse the health status of a person from the information collected on devices

**2. Analyzing Activity Patterns -**
Discovering the variables which help determine the kind of activity being done by a person

**3. Predict activity -**
Calculate a predictive model that can detect human activity from sensory detection signals

**4. Improve well-being -**
Creating individual exercise tables to improve a person's health

## 1.3 SUMMARY OF THE REPORT

We classified sensors for human action recognition into 3 broad categories - Vision sensors, Environment sensors, and RF sensors.

Vision sensors were further categorized into 3D cameras and depth cameras. Environment sensors consisted of both device sensors (smartphones) and body worn sensors (e-shoes, smartwatches, etc.). At last, RF sensors had their types as RFID and Wi-Fi sensors.

The main idea was to develop, design, and run a model that could recognize human activities and detect anomalies if any.

The goal of the project was to develop a Human Activity Recognition system using smartphone sensors, specifically the accelerometer, gyroscope, and magnetometer. The system aimed to accurately detect and classify the activities performed by a user using their smartphone. A custom raw dataset was collected and used for training and evaluation.
In the data preprocessing stage, the dataset was consolidated into a single file by concatenating multiple files collected from different individuals. The dataset was then split into training and validation subsets. Categorical labels in the dataset were encoded into numerical representations to facilitate machine learning algorithms. The continuous sensor data collected at regular intervals was converted into time series data to capture temporal dependencies and patterns.

The model architecture consisted of several layers. The first layer was an LSTM layer, which learned from a sequence of 100 data points at each timestamp and captured temporal dependencies. The output of the LSTM layer was passed to a Flatten layer to convert it into a 1D vector. A Dense layer followed, responsible for feature extraction, and a softmax layer served as the final classification layer, predicting probabilities for each activity.
The model was trained using the categorical cross-entropy loss measure and optimized using the Adam optimizer. Training was performed for 5 epochs, and the best model was saved using the Model Checkpoint callback.

# CHAPTER 2: DESCRIPTION

Standard PR methods have made great strides in HAR. However, there are a few drawbacks to the PR approach [22]. First, features are always rendered in a heuristic and hand-crafted way, based largely on personal knowledge or background information. This personal information may be helpful in certain job related settings, but in common areas and jobs, this will result in a lower and longer-term opportunity to build an effective job monitoring system. Second, only the shallow aspects can be learned from personal experience. Those shallow elements often refer to specific mathematical information including meaning, variation, frequency [12], magnitude, etc. They can only be used to identify low level activities such as walking or running, as well as activities that are difficult to say high-level or aware of context. For example, having coffee is complex and almost impossible to detect by using only shallow elements.

Third, conventional PR methods often require a large amount of well-labeled data in order to train the model. However, most work data remains labeled for actual applications [14]. Therefore, the effectiveness of these models is underestimated in unsupervised learning activities. In contrast, the existing deep production networks are able to exploit non-label samples of model training.

## 2.1 Deep Learning and Human Activity Recognition

In literature, there were two main methods used to develop a HAR model as mentioned above, which enable the free attachment of a phone to any on-body position. The first one is position independent HAR. In this approach, mixed sensors were used in building the model. This method trained the model using the data obtained from different positions. Additionally, some special handcrafted features were also used which restricted the variation in motion data in a different position for the same activity. Some domain experts suggested that the restriction of one activity in different positions is the limitation of handcrafted features.

Generalization with different settings for the same problem cannot be done as the features are shallow. The second approach is position aware HAR. It is based on building two or more classifiers. The first classifier is used to recognize the specific position of the sensor. Meanwhile, the second classifier is used to recognize the position-specific activity. the limitation of this method is that it is highly expensive to run on such a small device as a smartphone.

Another researcher [22] surveyed mobile and wearable sensor-based HAR. Categorization of deep learning methods was done based on generative, discriminative, and hybrid methods. The limitation of this survey was that it was restricted to only mobile and wearable sensor-based activity recognition. Whereas [Li et al.] another researcher presented different neural networks for radar-based activity recognition. His survey, however, was only limited to techniques such as CNNs and RNNs. The scope can be expanded to tackle specific challenges such as deep transfer learning and multimodal fusion.

In 2019, some researchers [16] used SVM based on micro-Doppler signatures for HAR. Features were extracted manually. A decision-tree structure was employed and a gesture recognition system was built using 60 GHz mm-wave radar. To perform real-time gesture recognition, a random forest classifier was employed in this system. An improved DTW algorithm was proposed for hand gesture recognition with a terahertz radar. It was capable of fully exploring the properties of range profile and Doppler signatures.

One very effective and most direct approach to Human Activity Recognition is to use Video based model. When talking about a video-based approach, naturally the camera comes into play. Cameras are used for recording videos and images to recognize participants' behavior.

Cameras provide rich and unique sets of information that are way better than other sensors. They offer continuous monitoring and intelligent processing throughout the process. This method provides ground truth, which can be used to check the results and improve the accuracy of machine learning in the real world.

## Classification Methods

Over the period, many authors introduced different methods of classification of Video-based HAR.
An author proposed a method which is Late fusion [12]. This model combines frames by linking the first and last frames in the clip. Early fusion was another method that was introduced which extracted all the local features from the same patches and locally linked them before encoding. Some other authors proposed more different methods like using CNN with LSTM, using Pose detection and LSTM, and
using Slow-fast Networks.

## Types of Activity Recognition Problems

There are basically three types of HAR problems. The first one is Simple Activity Recognition in which the model takes in a short video clip and further classifies the singular global action being performed.

The second one is Temporal Activity Recognition which takes a long video clipping containing multiple activities at different intervals of time. One of the two parts of this architecture localizes each individual action into temporal proposals. Whereas the second part focuses on classifying each video proposal.

Lastly, there is Spatio-Temporal Detection in which the video clipping contains multiple actions performed by multiple people. Each person is detected and localized in the video and the performed activities of everyone are classified.

## RFID

RFID stands for Radio Frequency Identification. RFID tags are placed at different positions on the participant's body. They have been used to track the vertical and horizontal positioning of a patient accurately in a confined space. RFID uses an analyzed dataset against a set of reference position datasets [18].

RFID is used in Trauma Resuscitation which is a quick-paced and highly dynamic process to treat severely wounded patients immediately after injury. RFID technology improved the rates of documentation and compliance of attending physician arrival to trauma activations.

In addition, many existing PR models are primarily focused on learning from static data; while job data in real life comes with streaming, it requires solid online learning and growing learning. Deep learning often overcomes those barriers. Features can be automatically read over the network instead of in person.

In addition, a deep neural network can also produce high-level representation in a deep layer [18, 21], making it more suitable for complex task recognition tasks. When faced with a large amount of unlabeled data, an in-depth productive model is able to exploit non-label training data [8, 20, 36]. In addition, in-depth learning models trained in a large label database can often be transferred to new jobs where there are few or no labels.

The deep learning model is famous for its ability to learn features through neural networks, receive large amounts of raw data for training, and identify intangible data through information transmission methods. Based on the Recurrent Neural Network (RNN) [23,31], the LSTM network model is known for performing well in extracting signal patterns in the input element space. By receiving input data that passes through a long sequence, it is specialized in time series [4] problems, especially in the timeline guess field.

The LSTM gate architecture changes memory by modeling local dependence on features. HAR is based on the assumption that dynamic sensory signals produce discriminatory patterns associated with different functions. Since such activities usually last for some time when they occur, HAR is considered to be a problem of time series and interdependence, in which the input data closest to the space may depend while the long sequence of samples at the time is considered. independent. Thus, the LSTM network model [15,16] is famous for its performance in HAR domains.

Human Activity Recognition (HAR) is done in different stages like collecting the data, preprocessing and training the data, and then finally recognizing the type of activity through various methods [8]. One common method is capturing the data through smartphones and wearable sensors. Wearable sensors are integrated into wearable devices or attached directly to the body to detect the kind of activity being performed by the user.

Sometimes, users might refrain from adapting to wearable sensor techniques. In such cases, smartphone sensors are taken into account. Device sensors capture data with the help of commonly used built-in sensors like accelerometer, magnetometer, and gyroscope on smartphone devices [Alexandre Bordat, Charissa Ann Ronao]

Smartphones and wearable devices are considered to be a good choice and consist of a large number of high-precision sensors that are integrated into these devices [17].

Many devices are present that can be used to capture different motions and activities performed by the user, but the methods or principles stay nearly the same.

1. Changes in the data with respect to different movements or actions performed by the user are recorded using sensors like an accelerometer and gyroscope
2. This data is then sent to desktops or other systems for computation using some complex algorithms
3. Data processing is done on the basis of the number of parameters which finally helps in analyzing and recording the kind of activity being done by the user [18].

Smartphones and wearable devices tend to offer many benefits as compared to other practices in case of Human Activity Recognition. No complex or expensive setup is required to predict the activity, built-in sensors are used to note down the motion [Shaohua Wan]. Triaxial or 3-axis accelerometers are generally used for detecting activities like walking, jogging, running, to name a few. [Hassan MM, Xiaolong Xu].

Built-in sensors are able to capture more continuous data and signals which helps in delivering accurate results [Anna Ferrari]. We can easily get information regarding the user's linear acceleration, velocity, direction and the electromagnetic radiation with the help of accelerometer, gyroscope and magnetometer sensors present in the smartphones and wearable devices

[Jeffrey W. Lockhart]. In addition, the user is always in the range of the sensor to capture continuous data, unlike in the case of video sensors [Hamid M. Ali].

Contrary to the advantages stated, some researchers feel that while smartphones and wearable devices are a fairly good choice for activity recognition, there can be certain limitations too [Peter James, Marcin Straczkiewicz]. Both smartphones and wearable devices need frequent charging or changing of batteries. Also, the user might not always have wearable device wrapped around the body, specially in case of an elder person as the user, but if used consistently, it can prove to be beneficial [Kevin Moore, Emma O'Shea].

Overall, smartphones and wearable devices are considered to be a much better choice as they maintain the user's privacy, provide continuous measurement of data, and produce more accurate results because of the presence of in-built sensors that work around a number of parameters in the 3d space [Md Zia Uddin, Lianyong Qi].

There can be broadly two ways of deploying wearable devices for HAR: 1. Using a 3-axis accelerometer or body area network (BSN) [Rahat Ali Khan] 2. Sensors used in combination with other sensors like temperature sensors, gyroscopes, etc. [Yan Wang]
With time, researchers have started to prefer smartphones as the ideal solution for the accurate recognition of activities. [Jukka-Pekka Onnela] This is because of the cognitive and computing power, exceptional processing capabilities, easy deployment option, and robustness of smartphones. [Joao Gama, Ms.S.Roobini]

Smartphones also consist of a variety of sensors like accelerometers, gyroscopes, and magnetometers and have wireless connectivity, which makes them very useful for purposes like smart home monitoring [Jing Zhao, Menghan Wu]. Smartphones consist of inertial sensors that use appropriate sensing resources to obtain HAR information.

Because of the presence of high-quality built-in sensors, wearable devices are automatically able to filter external magnetic interferences [Junhuai Li]. They are able to estimate the acceleration and angular velocity accurately [Shibo Zhang].

Sensor-based identification is much preferred because of its many benefits like small size, ease to carry, high sensitivity, and anti-interference ability. Sensors identify the physical states & characteristics using Inertial Measurement Units (IMU) sensors like accelerometers and gyroscopes, and magnetic field sensors which are altogether used to identify the activity [Ferhat Attal, Samer Mohammed].

Several studies, experiments, and research have been approved over the years to get precise and accurate results for Human Activity Recognition using Deep Learning algorithms, Machine Learning, etc., here are the details of some foremost research papers, thoroughly examined.

In research by Ronald Mutegeki [15], for Human Activity Recognition, a holistic deep learning-based activity recognition architecture, a convolutional neural network-long short-term memory network (CNN-LSTM) has been proposed. The idea of combining CNN with LSTM not only improves the predictive accuracy of human activities from raw data but also reduces the complexity of the model. Analysis of Human activity recognition (HAR) aims at using sensors; accelerometer, gyroscope, magnetometer, and others; that are built into IMU devices, and smartphones to recognize the activity being performed by the user of the device.

In research by Salwa O. Slim, studying human activity recognition shows that researchers are interested mostly in the daily activities of humans. The input of HAR models is the reading of the raw sensor data and the output is the prediction of the user's motion activities. The research shows that recently deep learning was used more than traditional machine learning, it also showed that CNN deep learning is mostly used; even though RNN and AE achieved a satisfying accuracy. The studies focus on recognizing the number of activities and different classification methods used for the recognition process.

Deep learning is a type of machine learning and artificial intelligence (AI) that imitates the way humans gain certain types of knowledge. Deep learning is an important element of data science, which includes statistics and predictive modeling.

**Convolutional Neural Networks (CNN)** are great for image data and **Long-Short Term Memory (LSTM)** networks are great when working with sequence data but when you combine both of them, you get the best of both worlds, and you solve difficult computer vision problems like HAR.

The CNN Long Short-Term Memory Network or CNN LSTM for short is an LSTM architecture specifically designed for sequence prediction problems with spatial inputs, like images or videos.

The CNN-LSTM model is both spatially and temporally deep and achieved better performance when it was compared with other deep learning approaches that use raw signal data as input.

## 2.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) [2] are the most popular method used for image analysis. Although image-type data is commonly used with CNN, it has the ability to process different types of data for analysis or segmentation. CNN is a neural artificial network that has some kind of expertise in selecting or finding patterns and making sense of these pattern acquisitions. CNN properties contain input, output, and hidden layers.

The Convolutional layers and combinations found in the hidden layer are layers of a special type of CNN. The convolutional layer converts acquired data into applications using convolution functions, so it throws the converted information into the next layer. This modification can be considered as the data is processed by different filters [31] and mapping features. A fully integrated layer also known as the Dense Layer acts as a traditional neural network, by connecting the neuron to the base of all the neurons in the next layer.

Convolutional neural network architecture is inspired by the hierarchical structure of the human visual cortex, which processes the visual information coming from the eye through a series of ordered and interconnected visual areas that perform feature recognition, fromsimple edge detection in the first areas to more complex shape structures in the higher levels.36 CNNs have gained popularity due to their ability to learn suitable representations and capture local dependencies from images or temporal series.

In the last few years, the use of deep CNN models has led to very good performance on a variety of problems, such as visual and speech recognition. Human activity recognition is also a good field for convolutional architectures, especially when considering the translation invariance and temporally correlated readings of time-series signals from activities, and their hierarchical structure as a combination of small movements. Due to this potential to identify the representative patterns of HAR's signals, CNN have recently been applied to human activity recognition in many research articles.

The operation performed by a convolutional layer consists of an elementwise product followed by a sum. The input, as a 2D matrix, is convoluted with a learnable kernel, a 2D matrix of a particular size, in a sliding-window fashion. The result of this operation forms the output feature map, which is another 2D matrix.

Note that more than one kernel can be applied to the input, hence the output will be composed of asmany feature maps as kernels are used. Different kernels will perform different convolution operations on the inputs, such as edge detection or sharpening. Each kernel can be considered as a specific feature detector, so the key task when training a convolutional neural network is to get it to learn the best kernels, those that extract the most meaningful features from the input.

Alternatively, a modified layout of the line unit (ReLU) can also be found on CNN. The ReLU framework simply defines it as an indirect function of keeping the positive values unchanged and making negative values zero.

## 2.3 Long Short-Term Memory

The structure of the Short-Term Memory System (LSTM), in general, provides effective results for long-term data such as speech signals or time-series data. LSTM [15] is a form of Recurrent Neural Network (RNN). The difference between these two approaches is the memory cell in LSTM. The memory cell has a gate structure that contains input, output, and memory.

Through these gates, cells decide to remember or forget information. So the gates control whether the cell reads writes, or resumes. To define the weight and bias of the LSTM network, each gate has limits. Boundaries can be determined or updated at the training stage.

In a study by [3] it was stated that CNNs are popular deep learning models for computer vision. The structure of a CNN resembles the visual cortex in the human brain in many ways. CNNs can extract features (i.e., spatial and temporal relationships) and separate the objects in the input picture using various filters. Convolution layers are made up of filters, and some fully linked layers often follow that to do the classification operation.

In addition to being effective feature learners, CNNs can scale to enormous datasets by using some pooling layers. In actuality, the goal of layer pooling is to decrease the dimensionality of input data while also extracting dominant features that are rotation- and position-invariant. Additionally, pooling layers play a crucial role in extracting dominant features that are robust to rotation and position variations. The max pooling operation, for example, selects the most prominent features within each pooling region. Consequently, these dominant features persist even when they undergo slight shifts or rotations in the input data. This characteristic is highly desirable in many computer vision tasks as it enables the network to learn features that are invariant to these spatial transformations, making the model more robust and generalizable.

Pooling layers in CNNs serve the dual purpose of reducing dimensionality and extracting rotation- and position-invariant dominant features. They effectively decrease the complexity of the network, allowing it to scale to large datasets while also enhancing its ability to learn meaningful representations from the data. By combining dimensionality reduction and feature invariance, pooling layers contribute significantly to the success of CNNs in computer vision tasks.

[21] stated that due to the nature of its design, the long short-term memory (LSTM) is a time recurrent neural network that is excellent for modelling time

series data. With horizontal lines passing through the top of the graph, the cells' current states are the key to LSTM.
The status of the cell is comparable to a conveyor belt. With only a few linear interactions, it is directly applied to the entire chain. Keeping the information flowing on it is simple. Through a cleverly created structure known as a "gate," LSTM has the capacity to add or delete information from the state of the cell. A door is a method to pass with knowledge. They include a pointwise multiplication operation and a layer of sigmoid neural networks.

In HAR, various human activities such as walking, running, sitting, sleeping, standing, showering, cooking, driving, opening the door, abnormal activities, etc. are recognized. The data can be collected from wearable sensors [2,12,25] or accelerometers [14] or through video frames [35] or images. HAR can be extensively used in medical diagnosis. For keeping track of elderly people, HAR can be used.

Crime rates can be controlled using HAR by monitoring. The smart home environment can be created by daily activity recognition. Driving activities can be recognized and lead to safe travel. Military actions can be recognized using HAR.

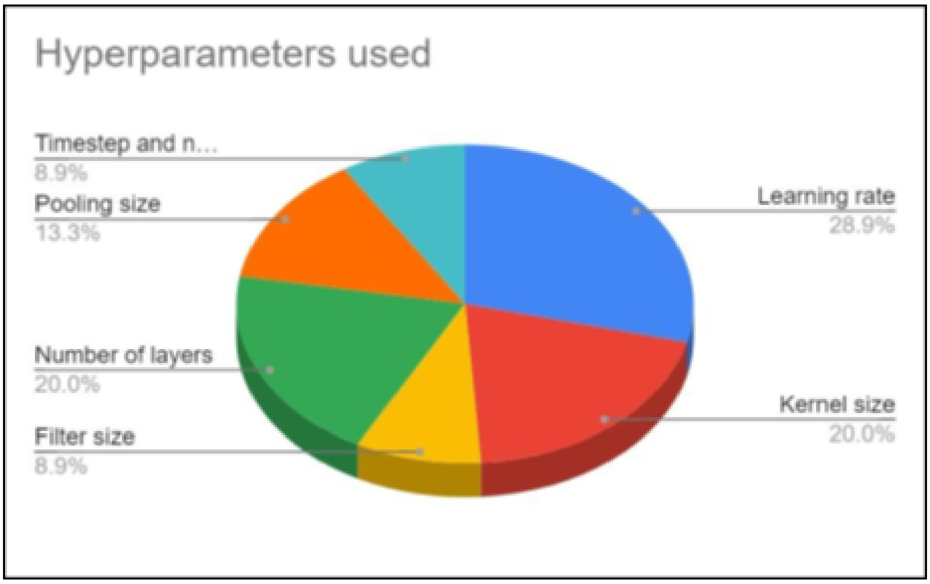


**Fig. 2.1 Pie chart showing types of hyperparameters taken for different papers**

The table and the pie chart above clearly show that learning rate is the parameter used in the maximum proposed models for Human Activity Recognition.

Timestep and neurons [4] in hidden layers along with the filter size have been used in the least number of papers that have proposed the model for Human Activity Recognition.

Another categorization has been done on the type of data source and sensors used in various proposed models. We have inferred that the environment sensors which include the device sensors and wearable sensors have been used most times for recognizing human activities

## Flowchart of the progress of the major project:

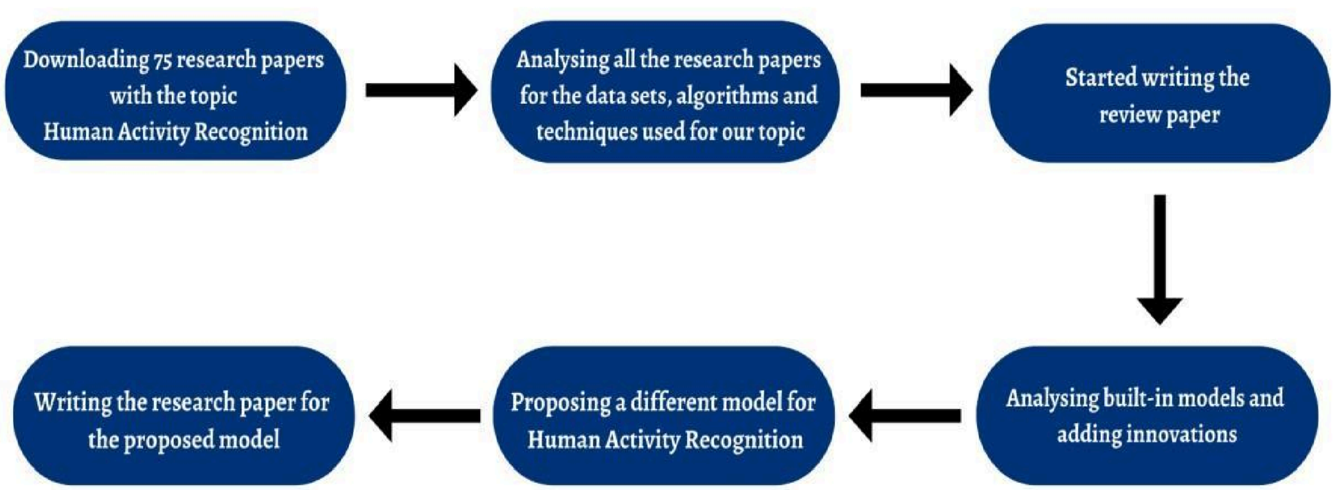


**Fig 2.2 Progress Flowchart for the major project**

We have studied 75 research papers and categorized them on various topics such as the sensors used, hyperparameters, datasets, and technology used.

# CHAPTER 3: AI TECHNIQUES IN OUR HAR MODEL

The main idea was to develop, design, and run a model that could recognize human activities and detect anomalies if any.
We built a Human Activity Recognition system with the help of smartphone sensors which include accelerometer, gyroscope and magnetometer, that plans to detect the kind of activity being performed by a user using the smartphone. A custom raw dataset has been used to detect the activities in our model.

Smartphone sensors are commonly used in human action recognition models due to their ubiquity and ability to capture relevant data for activity classification. The sensors typically used for this purpose include:

1. Accelerometer: Measures the acceleration experienced by the smartphone in three axes (X, Y, Z). It detects changes in linear motion and can capture movements such as walking, running, or jumping.

2. Gyroscope: Measures the angular velocity or rotational rate of the smartphone around its three axes. It helps detect and quantify rotational movements such as tilting or rotating the device.

3. Magnetometer: Measures the strength and direction of the magnetic field around the smartphone. It is used to detect orientation and heading changes, such as determining if the smartphone is being held in a portrait or landscape orientation.

4. Barometer: Measures changes in atmospheric pressure. It can be utilized to estimate changes in elevation, such as climbing up or down stairs.

5. GPS (Global Positioning System): Uses satellite signals to determine the smartphone's geographic location, allowing for the recognition of outdoor activities or tracking movement trajectories.

6. Proximity Sensor: Detects the presence of nearby objects or obstacles by emitting and receiving infrared or electromagnetic signals. It can be used to determine actions such as picking up or putting down the smartphone.

7. Ambient Light Sensor: Measures the intensity of ambient light. It can provide contextual information about the lighting conditions during specific activities, such as being in a bright or dark environment.

By collecting data from these sensors, it is possible to capture various aspects of human motion and activities. The sensor data can then be used as input for machine learning models, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), or hybrid models, to recognize and classify human actions accurately. These models can learn patterns and features from the sensor data to distinguish between different activities, enabling applications such as activity monitoring, fitness tracking, and gesture-based control systems.

The dataset used here includes sensor values from three different sensors: the Accelerometer, Linear Acceleration Sensor, and Gyroscope sensor. These sensors are typically found in devices like smartphones or wearable devices and are used to capture movement and orientation data.

For each timestamp in the dataset, there are corresponding sensor values recorded from these three sensors. These values represent the measurements captured by the sensors at that specific point in time. The specific sensor values might include data such as acceleration, linear acceleration, and angular velocity.

Additionally, the dataset includes activity labels that correspond to each data point or timestamp. These labels indicate the specific activity or behavior that was being performed when the sensor values were recorded. The activities recorded in this dataset include:

1. Walking
2. Standing
3. Jogging
4. Sitting
5. Biking
6. Going Upstairs

7. Going Downstairs

These activity labels provide valuable information about the context in which the sensor data was collected. They allow for the association of sensor measurements with specific activities, enabling the development of models that can recognize and classify different activities based on the sensor data.

Overall, this dataset presents an opportunity to explore and analyze sensor data captured from the Accelerometer, Linear Acceleration Sensor, and Gyroscope sensor, with corresponding activity labels that can be used for activity recognition or classification tasks.

In the data preprocessing stage, several steps are undertaken to prepare the dataset for machine learning tasks. Firstly, since the dataset was initially divided into 10 separate files, each collected by different individuals using their devices, these files are concatenated into a single file. This consolidation ensures that all the data from various sources is combined into one cohesive dataset.

Next, the dataset is split into two subsets: the training set and the validation set. This division allows for model training on the training set, while the validation set serves to evaluate the model's performance and fine-tune its parameters.

To facilitate machine learning algorithms that require numeric inputs, the categorical labels in the dataset are encoded into numerical representations. This step transforms the labels into a format that can be processed and utilized by the algorithms effectively.

Furthermore, as the data collected by sensors is continuous in nature and captured at regular intervals or timestamps, it is beneficial to convert it into time series data. This conversion enables the data to be represented as a sequence of values over time, making it compatible with models such as Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) models. These models excel at capturing temporal dependencies and patterns, which can be crucial in analysing time-dependent data.

By following these pre-processing steps, the dataset is prepared and structured in a way that optimizes its compatibility with subsequent machine learning tasks, such as model training, evaluation, and prediction.

The architecture of the model consists of several layers designed to process and classify sequential data.
Layer 1 is an LSTM layer, which stands for Long Short-Term Memory. This layer is responsible for learning from a sequence of 100 data points at each timestamp and returns a sequence mapping as output. LSTMs are a type of recurrent neural network (RNN) that are particularly effective in capturing and understanding temporal dependencies in data.

Layer 2 is a Flatten layer, which takes the 2-dimensional output from the previous LSTM layer (with dimensions representing the number of timestamps and the number of features) and converts it into a 1-dimensional vector. This flattening process is necessary to prepare the data for further classification.

Layer 3 marks the beginning of the classifier part of the model. It is a Dense layer, which means it is fully connected, and it takes the flattened output from the previous layer as input. The purpose of this layer is to transform the data and extract relevant features before passing it to the next layer.

Layer 4 is a softmax layer, which is the final layer of the model. It takes the input from Layer 3 and predicts the probabilities corresponding to each activity. Softmax activation is commonly used for multi-class classification problems, as it provides a normalized probability distribution over the classes.

The model utilizes Categorical Cross Entropy as the loss measure, which is a standard choice for multi-class classification tasks. It quantifies the difference between the predicted probabilities and the true labels.

The optimizer used for gradient descent is Adam or Adaptive Momentum. Adam is a popular optimization algorithm that adapts the learning rate based on the past gradients, allowing for efficient and effective parameter updates during training.

The model has been trained for 5 epochs, which means it has gone through the entire training dataset five times. The training process includes the use of a ModelCheckpoint callback, which saves the best model during training based on a specified metric, such as accuracy or validation loss. This ensures that the model with the best performance is saved and can be used later for inference or further analysis.

Most wearable devices such as smartphones [13,24] or clocks are equipped with an accelerometer, which is easy to find. The accelerometer has the ability to detect many types of everyday activities as most of them are simple body movements.

Nerve placement is also important. Most of the body-worn nerves are located on the wrist, waist, and hip joints. This placement strategy can help identify common daily activities [3,14].

However, when it comes to the opposing and ambient senses, it is important to use them in a non-invasive way. Those sensors usually do not communicate directly with users, so it is important to collect data naturally and without attack.

Before using in-depth models, raw sensor data needs to be processed in advance accordingly. There are two key factors. The first feature is a smooth window. Inputs should be cut into individual inputs according to sample size. This process is similar to standard PR methods.

Different sensory pathways can be treated as separate channels, and each sensor axis can be a channel. Using multiple channels can enhance the representation power of a deeper model as it can reveal hidden sensory information.

We also conducted studies on different datasets (which dataset was preferred more), various hyperparameters [23,36] that were taken into account, which method of deep learning was preferred, etc., and made statistical pie charts and bar graphs for the same. The engineering concept was discussed in AI about architecture, job losses, and development strategies.

Various modern results have been achieved by researchers in a simple way and atomic functions such as: running, walking, and stairs up, and down. But very little work is done with complex tasks where the work involves two or more simple actions.

For example, monitoring exercise when exercise such as burpees involves jumping, bending, and stretching the legs. That kind of complex and dangerous work requires attention to proper posture, but to the best of our knowledge, no HAR model can monitor physical activity that involves complex activity.

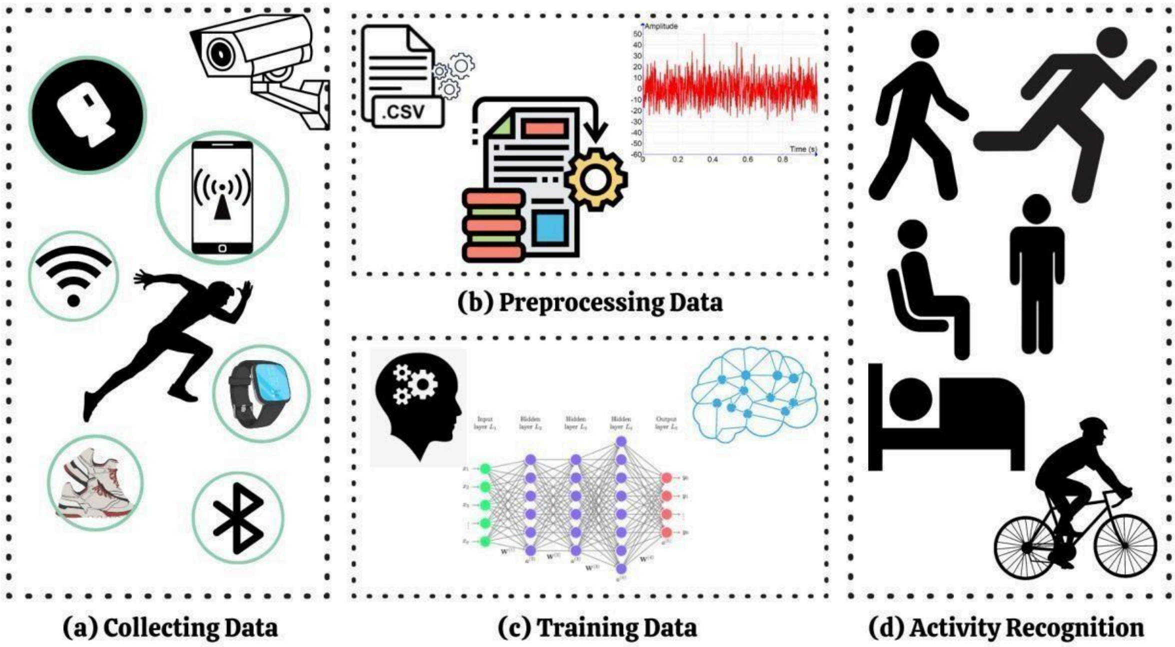


Fig 3.1 Process of creating the model for HAR

First the required dependencies are imported from Keras, including the Sequential model, Dense and Flatten layers, LSTM layer, L2 regularization, and the Adam optimizer.

A Sequential model is created, which allows for the sequential addition of layers.

The model architecture consists of several layers:

- An LSTM layer with 32 units, designed to capture temporal dependencies in the input data. It returns sequences instead of a single output and is regularized with L2 regularization on the kernel and bias weights.

- A Flatten layer is added to the model. This layer transforms the output from the previous LSTM layer into a 1D vector, which is required for the subsequent Dense layers.
- A Dense layer with 64 units and ReLU activation is added to the model. This layer performs feature extraction on the flattened input. L2 regularization is applied to the kernel and bias weights of the Dense layer.
- The final Dense layer has a number of units equal to the number of unique classes in the target variable. It uses the softmax activation function to generate class probabilities. L2 regularization is again applied to the kernel and bias weights.

In summary, the code constructs a Sequential model for Human Action Recognition. The model includes an LSTM layer to capture temporal dependencies, a Flatten layer to reshape the output, a Dense layer for feature extraction, and a final Dense layer for multi-class classification. L2 regularization is applied to the weights of all layers to prevent overfitting.

Various neural network architectures represent multifaceted functions. For instance, Convolutional neural networks (CNNs) can capture multimodal sensory input locally and the local translation invariance is accurate.

Deep neural networking can be detachable and scalable into interconnected networks with a global optimization feature that promotes various deep learning strategies learning.

We can define a CNN LSTM model in Keras by first defining the CNN layer or layers, wrapping them in a TimeDistributed layer and then defining the LSTM and output layers.

LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) architecture that has shown great effectiveness in human activity recognition using smartphone sensor data. This approach leverages the continuous and sequential nature of sensor measurements to capture temporal dependencies and patterns in the data.

In the context of human activity recognition, smartphone sensors like accelerometers, gyroscopes, and linear acceleration sensors provide information about the user's movements and actions. LSTM models are well-suited for processing such time series data, as they can effectively handle long-term dependencies and account for variations and complexities in the sensor readings.

The LSTM model for human activity recognition typically takes a sequence of sensor readings over a fixed time window as input. Each sensor reading represents a data point with multiple features. The LSTM layer in the model processes the sequential input, learning and remembering patterns across different time steps. By capturing temporal dependencies, the LSTM model can discern the underlying activities performed by the user.

The output layer of the LSTM model is typically a softmax layer that assigns probabilities to different activity classes, such as walking, running, or sitting. During training, the model is optimized to minimize the difference between the predicted activity probabilities and the ground truth labels.

By utilizing LSTM models for human activity recognition with smartphone sensors, it becomes possible to accurately classify and recognize various activities based on the sensor data. This has applications in fields like fitness tracking, healthcare monitoring, and context-aware applications, enabling personalized and adaptive experiences for users. LSTM's ability to capture long-term dependencies makes it particularly suitable for this task, as it can model the sequential nature of the sensor readings and extract meaningful features for accurate activity recognition.

# CHAPTER 4: RESULT

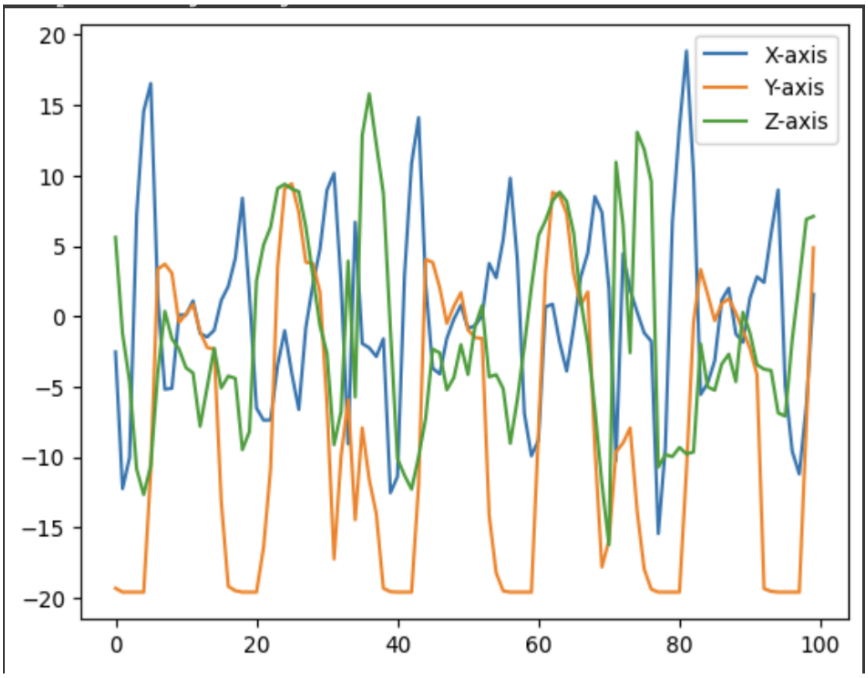


Figure 4.1 - Change in values while jogging

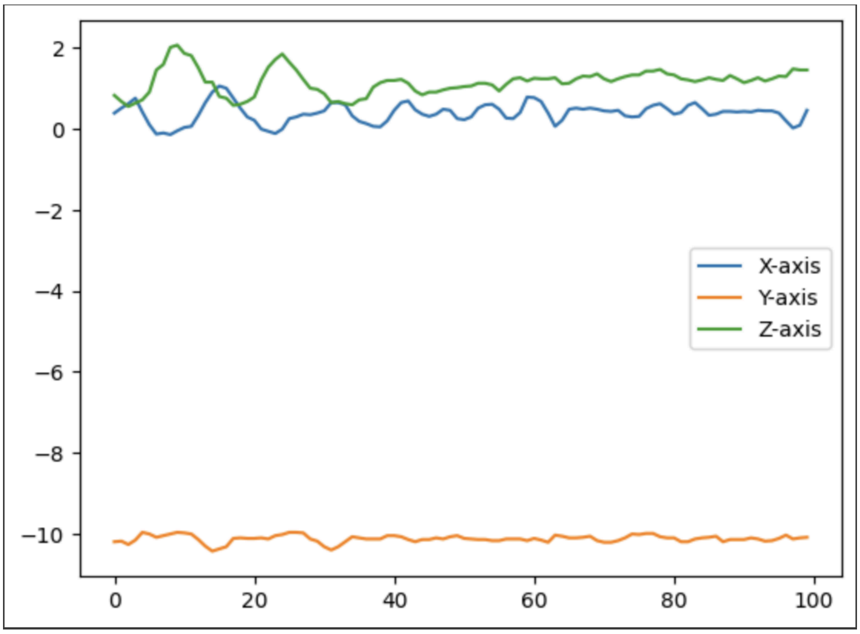


Figure 4.2 - Change in values while standing

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 LSTM_1 (LSTM)               (None, 100, 32)           5760

 Flatten (Flatten)           (None, 3200)              0

 Dense_1 (Dense)             (None, 32)                102432

 Dense_2 (Dense)             (None, 7)                 231

=================================================================
Total params: 108,423
Trainable params: 108,423
Non-trainable params: 0
_________________________________________________________________
```

figure 4.3 - Model Summary

# CHAPTER 5: CONCLUSION

Human activity recognition is an important research topic in pattern recognition and computing everywhere [1]. In this report, we explore the latest developments in in-depth learning methods for vision based Human activity recognition. Compared with conventional pattern recognition methods, in-depth reading minimizes dependence on the output of a man-made feature and gains better performance by automatically learning high-resolution representations of sensory data. We highlight the latest developments in three key areas: AI, applications [27,30], and sensors. After that, we summarize and discuss in more detail the survey research. Finally, a few major challenges and possible solutions are presented in future research

We suggested research around three HAR pillars namely, HAR devices, AI, and application domains. In the report, we hypothesized that growth in HAR devices is aligned with the evolving AI framework [22], and research corrects this by providing evidence of a clear representation of existing HAR models. Our second hypothesis states that AI growth is a HAR context that makes it suitable for domains with multiple characteristics. We measured this by introducing CNN [3] and TL [6] models for HAR models and also discussed the importance of hyperparameters, settings, and loss functions in the construction of HAR models. A unique contribution lies in the role of the AI framework in existing HAR models for each HAR device. The study also developed (1) HAR-based sensors with smaller devices [14,21] that would indicate the ground for healthcare application opportunities, particularly remote care, and monitoring, and (2) free HAR using a viable Wi-Fi device [23,28,32]. the use of HAR as an integral part of a healthy human life.

Finally, we built a Human Activity Recognition application using smartphone sensors, that plans to detect the kind of activity being performed by a user based on the sensor readings.

# REFERENCES


[1] Zhiwen Xiao, Xin Xu, A federated learning system with enhanced feature extraction for human activity recognition, Elsevier, 28 July 2021

[2] Sravan Kumar Challa, Akhilesh Kumar, A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data, Springer, 5 August 2021

[3] Xiaoli Li, Min Wu, Deep Learning for Human Activity Recognition, Springer, 8 January 2021

[4] Luay Alawneh, Tamam Alsarhan, Enhancing human activity recognition using deep learning and time series augmented data, Springer, 16 December 2020

[5] Xinyu Li, Yuan He, Semisupervised Human Activity Recognition With Radar MicroDoppler Signatures, IEEE, 16 August 2021

[6] S. Pushpalatha, Shrishail Math, Hybrid deep learning framework for human activity recognition, NAA, September 2021

[7] Farzan M. Noori, MD. Zia Uddin, Ultra-Wideband Radar-Based Activity Recognition Using Deep Learning, IEEE, 4 October 2021

[8] Sylvain Iloga, Alexandre Bordat, Human Activity Recognition Based on Acceleration Data From Smartphones Using HMMs, IEEE, 4 October 2021

[9] Md. Zia Uddin, Mohammad Mehedi Hassan, Activity Recognition for Cognitive Assistance Using Body Sensors Data and Deep Convolutional Neural Network, IEEE, 2018

[10] Junhua Ye, Xin Li, Deep Learning-Based Human Activity Real-Time Recognition for Pedestrian Navigation, MDPI, 30 April 2020

[11] Cagatay Berke Erdas, Selda Guney, HAR by using different deep learning approaches for wearable sensors, Springer, 2 February 2021

[12] Julien Maitre, Kevin Bouchard, Alternative deep learning architectures for feature level fusion in HAR, Springer, 8 March 2021

[13] Pranjal Kumar, Siddhartha Chauhan, HAR with deep learning: Overview, Challenges, and Possibilities, Springer, 27 March 2021

[14] Sakorn Mekruksavanich, Anuchit Jitpattanakul, Biometric user identification based on

HAR, MDPI, 27 January 2021

[15] Ronald Mutegeki, Dong Seog, Han A CNN-LSTM Approach to Human Activity

Recognition, IEEE, 2020

[16] E. Ramanujam, Thinagaran Perumal, Human Activity Recognition With Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review, IEEE, 2021

[17] Sakorn Mekruksavanich, Anuchit Jitpattanakul, Smartwatch-based Human Activity Recognition Using Hybrid LSTM Network, IEEE, 2020

[18] Pritam Khan, Bathula Shiva Karthik Reddy, Differential Channel State Information based Human Activity Recognition in IoT Networks, IEEE, 2020

[19] Ömer Faruk İnce, Ibrahim Furkan Ince, Human activity recognition with analysis of angles between skeletal joints using a RGB‐depth sensor, Wiley, 2019

[20] Wafa Shuaieb, George Oguntala, RFID RSS Fingerprinting System for Wearable Human Activity Recognition, MDPI, 2020

[21] Cuong Pham, Nguyen Ngoc Diep, e-Shoes: Smart Shoes for Unobtrusive Human Activity Recognition, IEEE, 2017

[22] Cuong Pham, Son Nguyen-Thai, SensCapsNet: Deep Neural Network for Non-Obtrusive Sensing Based Human Activity Recognition, IEEE, 2020

[23] Huan Yan, Yong Zhang, WiAct: A Passive WiFi-based Human Activity Recognition System, IEEE, 2019

[24] Kang Li, Xiaoguang Zhao, Sequential Learning for Multimodal 3D Human Activity Recognition with Long-Short Term Memory, IEEE, 2017

[25] Md. Zia Uddin, Jim Torresen, A Deep Learning-Based Human Activity Recognition in Darkness, IEEE, 2018

[26] Sumaira Ghaza, Umar S. Khan, Human activity recognition using 2D skeleton data and supervised machine learning, IET, 2019

[27] Salwa O. Slim, Ayman Atia, Human Activity Recognition based on Acceleration Data, IJACSA, 2019

[28] Zhenghua Chen, Le Zhang, WiFi CSI Based Passive Human Activity Recognition Using Attention Based BLSTM, IEEE, 2018

[29] Faeghe Fereidoonian, Farshad Firouzi, Human Activity Recognition: From Sensors to Applications, IEEE, 2020

[30] Türker Tuncer, Fatih Ertam, Novel tent pooling based human activity recognition approach, Springer, 2020

[31] Weiqing Huang, Yi Liu, TSCNN: A 3D Convolutional Activity Recognition Network Based on RFID RSSI, IEEE, 2020

[32] Yanwen Wang, Yuanqing Zheng, Modeling RFID Signal Reflection for Contact-free Activity Recognition, ACM, 2018

[33] Yegang Du, Yuto Lim, A Novel Human Activity Recognition and Prediction in Smart Home Based on Interaction, MDPI, 2019

[34] Zawar Hussain, Quan Z. Sheng, A review and categorization of techniques on devicefree human activity recognition, Elsevier 2020

[35] Luay Alawneh, Tamam Alsarhan, Enhancing human activity recognition using deep learning and time series augmented data, Springer, 2021

[36] Nurul Amin Choudhury, Soumen Moulik, Physique-based Human Activity Recognition using Ensemble Learning and Smartphone Sensors, IEEE, 2021

[37] Zhiwen Xiao, Xin Xu, A federated learning system with enhanced feature extraction for human activity recognition, Elsevier, 2021

[38] Abdul Rehman Javed, Raza Faheem, A smartphone sensors-based personalized human activity recognition system for sustainable smart cities, Elsevier 2021

[39] Emilio Sansano, Raúl Montoliu, A study of deep neural networks for human activity recognition, Wiley 2020

[40] Ling Chen, Xiaoze Liu, Deep learning based multimodal complex human activity recognition using wearable devices, Springer 2020