FUTURE REPRESENTTION MODEL FOR HUMAN ACTIVITY RECOGNITION USING CNN

Racha Revathi [1], Dr. M.M.Yamuna Devi [2]

[1]Research Scholar, *Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation ,Guntur-522302, India*

[2]*Associate ProfessorDepartment of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation ,Guntur-522302, India*

rrevathics@gmail.com,yamunadevimm@kluniversity.in

ABSTRACT:

Human activity recognition (HAR) research has become increasingly important to healthcare systems recently. The HAR's accurate activity classification results have wide-ranging applications that improve the performance of the healthcare system. The system forecasts abnormal actions based on user movements, and the HAR findings are helpful in keeping track of a person's health. The abnormal activity forecasts made by the HAR system improve healthcare monitoring and decrease medical issues among users. To achieve the best possible result, multiple machine learning and deep learning models have been developed for this area. As more effective and powerful deep learning approaches are developed, the employment of older, more conventional machine learning algorithms has declined in popularity. Convolution neural networks (CNN), long short-term memory (LSTM), recurrent neural networks (RNN), and other deep learning methods are frequently used. On numerous public datasets, the developed algorithm has been evaluated to show its efficacy, and in certain cases, it has outperformed state-of-the-art performance. Some tools are presented and discussed to ensure the suggested algorithm's effectiveness for continuous real-time human action recognition and to verify the algorithm's variability. In this article we will use hybrid model to improve performance of deep convolution neural networks (ConvNets) to detect human activities.

Keywords: HRM system, Deep Learning, LSTM, RNN,

INTRODUCTION:

Researchers have been very interested in human activity recognition in recent years because it has so many uses in everyday life, such as in healthcare, motion analysis, smart monitoring systems, and smart homes. HAR works on analyzing the data collected about how people act to understand and predict certain patterns of behavior. Different methods, such as accelerometers, infrared, RFID, and video, are used to gather the data. By encoding human behavior in different ways, different data sets give us different values and sources of information. Based on how the data was collected, HAR is now usually split into two groups: video-based and sensor-based. Video-based systems mostly use cameras to record moving pictures and images so that computer vision methods can be used to figure out how people act and what they do every day. Even though they have had some great results, they are affected by things like how the target is covered and how bright the light is. Sensors are built into many smart products, like smart phones and smart watches. Because smart gadgets are so common and useful in our daily lives, and because they can be carried around and do a lot of work, sensor-based systems have become an important area of HAR study.

Human activity recognition can be used in healthcare to keep track of what people with long-term diseases like Parkinson's disease do every day. This can make it easier for doctors and nurses to notice changes in a patient's behavior or ability to move around and make changes to treatment plans as needed. In sports, HAR can be used to evaluate an athlete's ability and technique, point out places where they could improve, and make training programs better. In smart homes, HAR can be used to automate appliances and gadgets based on what people do. For example, when someone walks into a room, the lights could turn on. In robotics, HAR can be used to improve robot skills, making it easier for them to interact with people and understand how they act. In terms of security, HAR can be used to find suspicious behavior or break-ins in places like banks, airports, and military bases. Overall, HAR is an area of study and development that is growing quickly and has many uses in many different fields and businesses. Different methods have been looked into in the past to try to solve the problem of recognizing human activities. Some of them are based on data from sensors, and others are based on data from video

[1]. Approaches that use sensor data usually require the person being shot to wear some kind of wearable sensor. In places like the home, it might not always be possible to have a sensor, which means that this method might not always be easy to use. Another way is to use a camera to take a video and then use the video to figure out what was going on. This plan might work better in cases where wearing a sensor would be hard. Some studies have looked into how to solve the problem of figuring out what people are doing in movies. One of these techniques is based on a hierarchical codebook model of local spatiotemporal video volumes [2]. This method, which is based on a bag of video words (BOV) representation, can be used even if you don't know anything about actions, background subtraction, or motion predictions. But a big problem with this method is that it can't be used to figure out long-term behaviors, like those that involve a number of actions that happen one after the other. Other methods, like Hidden Markov Models (HMM) [4] and Support Vector Machines (SVM) [3], have also been looked into and tested to see how well they can spot human behavior.

HAR systems can be used in a variety of applications, such as wellness, sports performance, athletics, healthcare, security, etc. They are typically monitored or unmonitored. The goal of the HAR system's modeling process is to predict the label of a person's behaviour from a still image or moving image, which is often accomplished by both image- and video-based activity recognition.

Other HAR problems include:

- inconsistencies in sensor data resulting from device location;
- movement variance;
- interference from concurrent operations;
- Noisy data that results in distortions; and time-consuming and expensive data collection techniques.

RELATED WORKS:

Annapoorani Subramanian, Sheryl Mathew, and others, 2023There are several uses for human activity recognition (HAR) in fields like healthcare, sports, and security. Numerous methods have been investigated in the past to approach the problem of identifying human activity due to its numerous applications. Deep learning, support vector machines, decision trees, and other machine learning methods have all been created to precisely identify human behaviours from sensor data. The ultimate goal of HAR is to enhance people's quality of life by making it accessible to them to get customized and preventative treatments based on patterns of human behavior. Jin Wang, et al Lee (2014) [5] explain how HAR was used in monitoring systems in common areas like banks and airports to stop thefts and other dangerous activities from occurring.. The results showed that the suggested methodologies are capable of early detection of continuing interactions between two or more people. Abdulhamit Subasi, et al (2019) [1] applying machine learning methods. HAR is often used to keep track of what older patients in rehabilitation facilities are doing to help treat and avoid chronic diseases. In order to monitor the everyday activities of elderly people, HAR is also included into smart homes[19]. In addition, HAR supports exergaming in rehabilitation facilities for patients with dysfunction and psychomotor slowness, post-stroke motor patients, and youngsters with motor disabilities. Djamila Romaissa Beddiar, et al 2020 indicate that human activity recognition has been used frequently in activities, exergaming, and for people with neurological injuries. [6] who presented a "Vision-based human activity recognition: a survey". Through HAR, human body motions are recognised to give the computer instructions on how to carry out specific jobs. Elderly folks and those with neurological injuries can readily interact with games and exercise games by using a simple gesture. By adopting standardised free-hand movements, HAR also enables surgeons to manage the intraoperative visual monitor virtually. Using computer and machine vision technology, human activity recognition, or HAR, interprets human movements. It has drawn a lot of interest because of its utility and adaptability, even if its full potential has not yet been explored. "A Review on Video-Based Human Activity Recognition" was offered by Shian-Ru Ke, Hoang Le Uyen Thuc,et al in 2013 [7]. Any approach is made up of four basic steps: gathering training data from external or wearable devices, reducing noise with windowing, getting structural and statistical features from the data, and classifying the data using machine learning and deep learning models. Vision-based and sensor-based are the two main types of human activity recognition, which are further broken down into wireless, object-tagged, and dense sensing. Homay Danaei Mehr et al (2019, p. 8) made a paper called "Human Activity

Recognition in Smart Home with Deep Learning Approach" in which human activity recognition is roughly put into these groups. In their paper, "Deep Learning-based Multimodal Complex Human Activity Recognition Using Wearable Devices," published in 2020 [9], et al provided an overview of previous research that uses a vision-based approach for activity recognition. They divided the literature into two main groups: unimodal and multi-modal approaches. The different kinds of unimodal methods are random, rule-based, space-time-based, and shape-based. Methods that are unimodal only use input from one modality. Multimodal methods, which use information from many different places, can be broken down even further into behavioral, efficient, and social networking approaches. " Ong Chin Ann et al (2015) [10] explains that people have been researching how to use technology for monitoring to figure out what people are doing for the last ten years. Various sensors are some of the most common types of sensors that are used to track behavior. Radio technology, like RFID, is used by some devices. Djamila Romaissa Beddiar, et al (2020) [6] suggests that machine learning helps activity recognition a lot. Once the information has been collected using a variety of methods and technologies, it is the task of a method that uses machine learning to infer or recognize the action. Salvatore Gaglio, et al in 2015 [11] describes the informtion about the process, some of the most popular machine learning methods are SVM, KNN, Random Forest, Naive Bayes, and HMM. Before putting a machine learning program to use, feature selection is another important step. It's possible that better results will come from a powerful mix of factors. Shugang Zhang, et al 2020[12] describes the information about the According to the methods used, HAR can also be split into three groups: action-based, interaction-based, and motion-based. Action-based technology is further broken down into gesture recognition, posture recognition (like standing, sitting, running, or cooking), behavior recognition, fall detection (like when a person's body position changes from normal to reclining without them realizing it), activities of daily living (like eating, sleeping, and drinking), and ambient assisted living. Adrian Nunez-Marcos, et al (2017) [13] describes the that motion-based activities are important for finding out if a person is there or not, which is very helpful for security and surveillance. RFID is the leader in this area, and the solutions are very accurate and are not excessively. There are various types of work in this region, such as tracking, detecting motion, and counting people. After different methods and technologies are used to collect data, it is up to an algorithm that uses machine learning to guess or recognize the action. Liandro B. Marinho, et al (2017) [3] describes the information which uses machine learning in a way that makes it easier to identify activities. In 2016, S.U. Park, J. H. Park,et [14]. al describes the Some of the most common machine learning algorithms for HAR . HAR can be used for healthcare for the elderly, smart settings, security and surveillance, human-computer interaction (HCI), indoor navigation, shopping experiences, and many other things. As we read through all the information, we find out that there were a lot of problems. The first group could include complex activities, such as exercises that combine simple actions like sitting, standing, etc., the presence of many subjects, and concurrent activities that combine several simple actions at once. The second thing is background noise and the effect of the surroundings. The fact that there are things with almost human-like shapes, cluttered backgrounds, and different levels of light intensity is a big problem that needs to be solved. The third thing is the security of the whole system. Most study is focused on accuracy, cost, and scalability, which makes the issue of security less important. Fourth, there aren't enough training samples to start with. Yang Wang and Greg Mori wrote in their 2009 article " [15] describes that any problem statement and potential solution needs a well-processed dataset that includes the range of actions needed. Also, most of the easily accessible datasets are either animated text or faces of people from other countries, both of which are hard for an Indian setting to process. With this as our main goal, we're trying to make a real dataset with movies that mostly show Indian faces and bodies, and we're using different suggested methods to evaluate and compare how accurate they are. Some of these things that people do around the house are standing, sitting, sneezing, and crying.

Proposed HAR system

To automatically recognize and categories human behaviors based on sensor data is the goal of Human Activity Recognition (HAR), an area of computer science and engineering. It's the use of sensors to identify human activity or motion based on gestures or motions of the body.

Depending on the architecture, HAR systems can be used to improve health, athletic performance, medical outcomes, security, and more.

The HAR system's goal is to model video-based activity recognition and image-based activity recognition to predict the label of a person's behaviour from a video or still image.

One of the most common applications of vision-based HAR systems is in the evaluation of poses. More and more studies are using it because it provides crucial insights into human behaviour. Useful for a wide variety of applications involving semantic comprehension, content extraction, and HAR. Different DL methods are used, but in particular convolutional neural networks.

Human anatomy, cultural signifiers, orientation, and pose variety present a significant barrier for HAR. As a case study, consider the image below. It could be difficult to tell if the individual is falling or trying to do a handstand. Because of this doubt, more recent approaches within the context of artificial intelligence are encouraged.

Incorporating increasingly complex features, utilising numerous data sources, and capturing the spatial and temporal interactions between body parts are all ways in which multi-modal learning and graph-based learning hope to increase the accuracy and resilience of HAR systems.

Data collection

Kinetics-700 Approximately 650000 Youtube video clips representing 700 distinct human action types are included in this high-quality video collection. Both human-object and human-human interactions are featured in the videos. Kinetics is a good dataset for training models to recognise human actions.More data sets include WISDM, HAR, and PAMAP2 from UCI.

The 50 action categories that make up YouTube's UCF50 dataset are all based in reality [16]. The dataset is difficult because of all the many camera movements, object poses, viewpoints, item sizes, backgrounds, lighting conditions, and so on. Dive, walk with a dog, jump rope and ride a horse are all activities that fall under this category. There are, on average, 133 videos in each action subcategory. We included walking, sitting, and leaping as the three most fundamental actions in our dataset. Our YouTube channel and free stock video sources provided the majority of the videos in our dataset, but we also included some footage of our friends and family members performing the activities. Due to the circumstances of its collection, the dataset largely features Indian faces. There are 21 videos in each activity category, ranging in length from 3 to 15 seconds.

Data pre-processing

Human activity recognition (HAR) relies heavily on the preprocessing stage, which involves cleaning, transforming, and preparing raw sensor data for subsequent analysis and modelling. Some standard preparation processes include:

2. **Filtering**Filtering is a method of signal processing used for removing undesirable signals and noise from raw sensor data. Common filters used in HAR for noise reduction and image improvement include low-pass filters, high-pass filters, and band-pass filters, all of which depend on the frequency range of the indications of interest.
3. **Feature extraction**: The nature of the task at hand and the available sensors will dictate which characteristics will be employed. Features like the mean and standard deviation can be extracted from accelerometer data, as can frequency-domain elements like the Fourier transform and wavelet transformation parameters.
4. **Feature selection:** Activity detection algorithms can be made more accurate and efficient by the feature-selection process, which reduces the dimensionality of the feature space. Selecting the most important features necessitates taking into account their exclusionary power, their relationship to the labelling of activities, and their redundancy with other features.
5. **Segmentation**: Segmentation necessitates partitioning the sensor data into smaller segments or windows in order to extract the temporal elements of the activity. The duration and significance of the phenomenon under observation define the size and overlap of the window. The data is then segmented, and the characteristics of each window are calculated.
6. **Normalization**: To ensure that features are comparable among sensors and people, features are scaled to have a neutral mean and variance of 1. This procedure is known as normalisation.

7. **Dimensionality reduction** Principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) reduce feature space dimensionality and remove redundant or irrelevant features.
8. **Missing Value Imputation**: Imputation involves completing gaps in sensor data. Data transmission errors or device malfunction could be to blame for the failure. For missing values, simple imputation methods can be used, such as mean or median interpolation.
9. 9Since it has an impact on the accuracy and dependability of activity identification models, data preparation is a vital step in HAR.

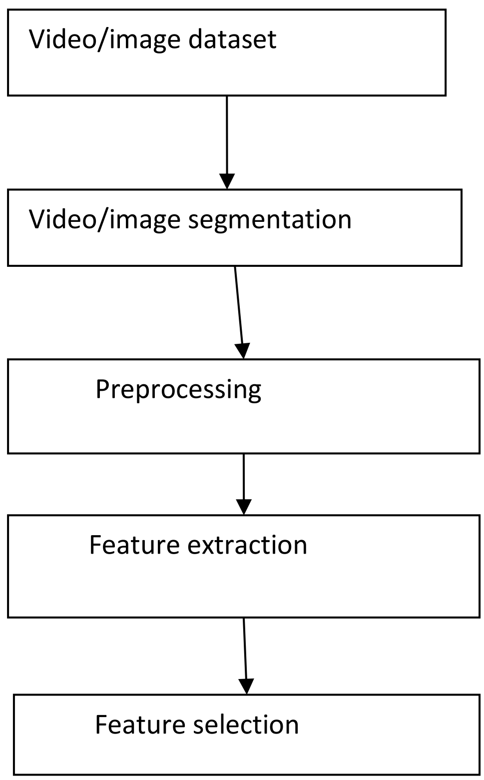

Figure 2: System Architecture

1. **Convolution Neural Networks :**CNNs are deep learning techniques that specialize in processing images and time-series data from sensors like gyroscopic and accelerometers to measure However, these algorithms may require more computing resources than other models and are prone to over fitting because of how well they manage hierarchical characteristics from raw data and complex data patterns.

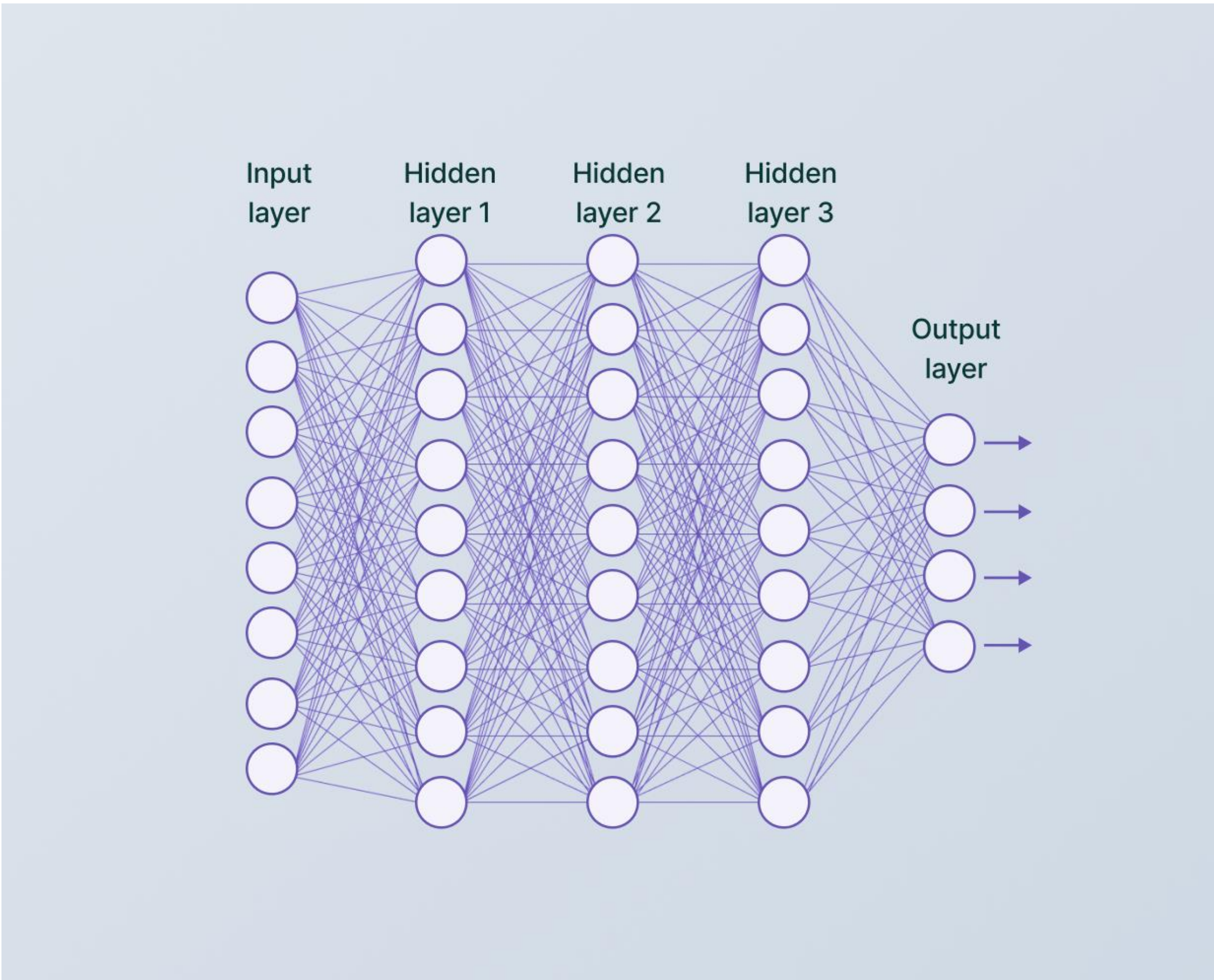

figure 2:CNN Architecture

Popular architectures for convolution networks

LeNet

This was the first time anyone used a convolution neural network. LeNet was taught to work with grayscale, 32x32x1-dimensional, 2D images. The goal was to find numbers written by hand on bank checks. It had two blocks of convolutional-pooling layers and then two layers that were fully linked.

AlexNet

The Imagenet dataset, which contains 15 million 256*256*3 high-resolution images, was used to train AlexNet.

For the first time, ReLU activation function and overlapping pooling with stride window size were utilised between convolution layers and pooling layers. It had three blocks of fully connected dense layers for classification, followed by five convolutional-pooling layer blocks.

Instead of continuously adding growing dense layers to the model, VGGNet VGGNet offered a technique to enhance performance.

The basic innovation consisted of arranging layers into blocks that were regularly used in the design since it was held that more layers of narrow convolutions were more effective than fewer levels of larger convolutions.

A VGG-block contained several 3x3 convolutions that were spaced out by 1 in order to maintain the size of the output and input, followed by maximum pooling to reduce the resolution by half. The architecture consisted of three dense layers that were completely connected, followed by n VGG blocks.

GoogLeNet

This architecture's Inception blocks are made up of 1x1x3x5 convolution layers, 3x3 max pooling, padding on the layer before it to make the result have the same structure as the inputs, and concatenating their outcome.

There are a total of 22 layers, none of which are fully connected. It employs 12 times less parameters than older models like AlexNet, although having 4 million overall.

ResNet

It was demonstrated that accuracy eventually reached saturation and started to degrade as network depth rose. Data scientists thus proposed skip connections as a fix. These connections make training quick, enable skipping one or more layers, and offer an alternative path for data and gradients to flow. On the premise that deeper models shouldn't cause more training error than their shallow equivalents, the concept of residual blocks was put out.

In reality, by designating other levels in the deeper network to be identity mapping, a deeper network was created from shallow networks.

DenseNet

Vanishing gradients were a ResNet flaw that was observed. The key to training deep networks was to make short paths from earlier layers to subsequent layers. Direct connections existed between each tier.

ZFNet

It is an adjustment to AlexNet. The main distinction is that AlexNet utilises 11x11 filters instead of 7x7 filters in its architecture since it believes that larger filters may result in information loss. These adjustments increased effectiveness and efficiency.

Short-Term Long Memory

A Recurrent Neural Network (RNN) variant known as Long Short-Term Memory (LSTM) has been successfully applied to a number of sequential data-related applications, including Human Activity Recognition (HAR).

ther RNNs, LSTM models are made to analyse data sequences and store internal memories of earlier inputs. This allows them to maintain the temporal relationships between various sequence segments.

The main advantage of LSTMs above all other RNNs is their ability to consciously forget or keep data from earlier time steps. In ordinary RNNs, vanishing gradients are a problem that this helps to resolve. In the input sequence, LSTMs can successfully replicate long-term dependence. They work well for challenging HAR tasks like detecting anomalies and distinguishing intricate human activities.

In numerous benchmark datasets, LSTM-based models showed significant improvements in HAR tasks, achieving state-of-the-art performance. They have additionally demonstrated tenacity in spotting intricate activities and coping with input sequences of varying duration. However, just like other deep learning models, LSTMs have a number of disadvantages for HAR, including the need for enormous amounts of labelled data, computational expense, and model interpretability.

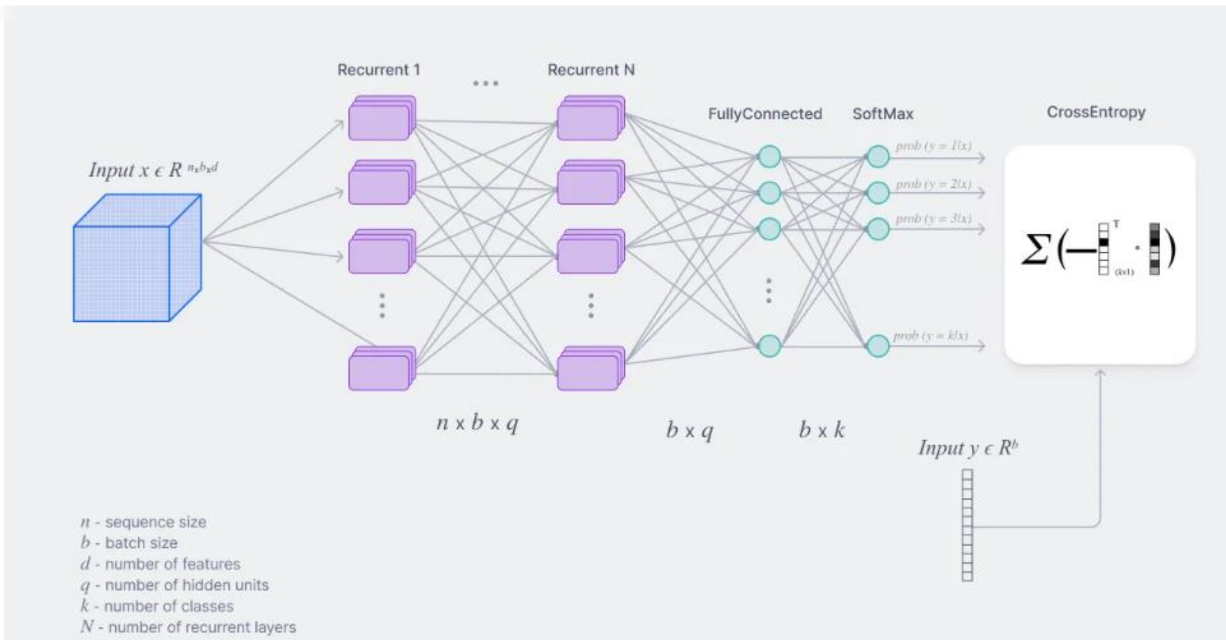

Figure 3:LSTM,RNN architecture

Conclusion:

HAR systems use sensor data to monitor human behavior and detect body motions and activities. Mobile and wearable devices' sensors enable most HAR applications. Using wearable device data, Deep Learning (DL) has expedited HAR research. This paper covers HAR, its applications, and public benchmark datasets. We also explored HAR-related DL methods and hybrid mrthodes to analyze performance in detecting activities from video or image . We also discussed the field's difficulties and potential directions for critical HAR research.

REFERENCES:

Chuanlin Zhang[1,3], Kai Cao[2,3], Limeng Lu[2,3] & Tao Deng[1,2,3*]. (2022). A multi-scale feature extraction fusion model for human activity recognition. *Scientific Reports*. 12(20620), p1-12.

1) Abdulhamit Subasi, Kholoud Khateeb, Tayeb Brahimi and Akila Sarirete , “Human activity recognition using machine learning methods in a smart healthcare environment” Version of Record 15, Nov 2019.

2) Mehrsan Javan Roshtkhari, Martin David Levine, “Human activity recognition in videos using a single example” Image and Vision Computing 31(11):864–876 , DOI:10.1016/j.imavis.2013.08.005, Nov 2013

3) Liandro B. Marinho, A.H. de souza junior and P.P.Reboucas Filho, “A New Approach to Human Activity Recognition Using Machine Learning Techniques” AISC,volume 557, Feb 2017

4) Neil Robertson and Ian Reid, “A general method for human activity recognition in video ” Sep 2006.

5) Jin Wang, Zhongqi Zhang, Li Bin and Sungyoung Lee ,“ An Enhanced Fall Detection System for Elderly Person Monitoring using Consumer Home Networks “ IEEE Transactions on Consumer Electronics 60(1):23-29 DOI:10.1109/TCE.2014.6780921, Feb 2014

6) Djamila Romaissa Beddiar, Brahim Nini, Mohammad Sabokrou and Abdenour Hadid “Vision-based human activity recognition: a survey” 79, pages30509–30555 (2020), Aug 2015

7) Shian-Ru Ke, Hoang Le Uyen Thuc, Young-Jin Lee, Jenq-Neng Hwang, Jang-Hee Yoo, Kyoung-Ho Choi ,“A Review on Video-Based Human Activity Recognition” Computers 2013, 2(2), 88-131; https://doi.org/10.3390/computers2020088 Jun 2013

8) Homay Danaei Mehr and Huseyin Polat “Human Activity Recognition in Smart Home With Deep Learning Approach” Aug 2019

9) Ling Chen, Xiaoze Liu, Liangying Peng and Menghan Wu "Deep learning based multimodal complex human activity recognition using wearable devices" 51, pages4029– 4042 Nov 2020

10) Ong Chin Ann and Bee Theng Lau "Human activity recognition: A review" DOI: 10.1109/ICCSCE.2014.7072750 Conference: Proceedings - 4th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2014, Mar 2015

11) Salvatore Gaglio, Giuseppe Lo Re and Marco Morana "Human Activity Recognition Process Using 3-D Posture Data" IEEE Transactions on Human-Machine Systems (Volume: 45, Issue: 5) Oct 2015 12) Shugang Zhang, Zhiqiang Wei, Jie Nie, Lei Huang, Shuang Wang and Zhen Li "A Review on Human Activity Recognition Using Vision-Based Method" Volume 2017 Article ID 3090343 , Aug 2017

13) Adrian Nunez-Marcos, Gorka Azkune and Ignacio Arganda-Carreras "Vision-Based Fall Detection with Convolutional Neural Networks" Wireless Communications and Mobile Computing 2017(1):1-16 DOI:10.1155/2017/9474806, Dec 2017

14) S.U.Park, J.H.Park, M.A.Al-Masni, M.A.Al-Antari, Md.Z.Uddin, T.S.Kim "A Depth Camera-based Human Activity Recognition via Deep Learning Recurrent Neural Network for Health and Social Care Services" Oct 2016

15) Yang Wang and Greg Mori "Human Action Recognition by Semi Latent Topic Models" IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 31, Issue: 10), Oct 2009

16)Chen, L., & Nugent, C. D. (2019). Human activity recognition and behaviour analysis. Springer International Publishing, http://dx.doi.org/10.1007/978-3-030-19408-6.

17)Fu, Y. (2016). Human activity recognition and the activities holter (p. VII, 174). Switzerland: Springer International Publishing, http://dx.doi.org/10.1007/978-3- 319-27004-3.

18)Sheryl Mathew, Annapoorani Subramanian, S. Pooja, Balamurugan MS. (2023). HUMAN ACTIVITY RECOGNITION USING DEEP LEARNING APPROACHES: SINGLE FRAME CNN AND CONVOLUTIONAL LSTM. www.researchgate.net publication.-(-), p.-.