GMR Institute of Technology
An Autonomous Institute Affiliated to JNTU-GV

HUMAN ACTIVITY RECOGNITION USING MACHINE LEARNING

A Mini Project Report Submitted in Partial Fulfillment for the Award of the Degree

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHNOLOGY

Submitted By

A.G.S.Anirudh	**(21341A1201)**
B.Jhansi	**(21341A1215)**
B.Sushma Gayathri	**(21341A1213)**
Akankhya Gantayat	**(21341A1202)**

Under the Esteemed Guidance of

Mrs. L. Swathi
Assistant Professor

May 2024

DEPARTMENT OF INFORMATION TECHNOLOGY

GMR INSTITUTE OF TECHNOLOGY

An Autonomous Institution Affiliated to JNTU-GV, Vizianagaram
(Accredited by NBA, NAAC with 'A' Grade & ISO 9001:2015
Certified Institution) GMR Nagar, Rajam-532 127, A.P

2023 - 24

GMR Institute of Technology
An Autonomous Institute Affiliated to JNTU-GV

GMR INSTITUTE OF TECHNOLOGY

(An Autonomous institute, affiliated to JNTU-GV, Vizianagaram)

NAAC "A" Graded, NBA Accredited, ISO 9001:2015 Certified Institution G.M.R. Nagar, Rajam-532127, A.P

DEPARTMENT OF INFORMATION TECHNOLOGY

CERTIFICATE

This is to certify that mini project report titled **"Human Activity Recognition Using Machine Learing"** *submitted by* **A.G.S Anirudh (21341A1201), B. Jhansi (21341A1215), B. Sushma Gayathri (21341A1213), Akankhya Ganatyat (21341A1202)** *has been carried out in partial fulfilment for the award of* **B.Tech.** *degree in the discipline of* **IT** *to* **JNTUGV** *is a record of bonafide work carried out under our guidance and supervision.*

The report embodied in this paper has not been submitted to any other university or institution for the award of any degree or diploma.

Signature of the Supervisor

Mrs .L. Swathi,
Assistant Professor **,**
Department of IT,
GMRIT,Rajam.

Signature of the H.O.D

Dr. V. Vasudha Rani
Associate Professor and HOD,
Department of IT,
GMRIT, Rajam.

ACKNOWLEDGEMENT

It gives us immense pleasure to express a deep sense of gratitude to our guide, **Mrs. L. Swathi,** Assistant Professor, Department of Information Technology for wholehearted and invaluable guidance throughout the report. Without her sustained and sincere effort, this report would not have taken this shape. She encouraged and helped us to overcome various difficulties that we have faced at various stages of my report.

We would like to sincerely thank **Dr. V.Vasudha Rani,** Associate Professor & HOD, Department of Information Technology, for providing all the necessary facilities that led to the successful completion of my report.

We take the privilege to thank our Principal **Dr. C. L. V. R. S. V. Prasad** who has made the atmosphere so easy to work. I shall always be indebted to them.

We would like to thank all the faculty members of the Department of Information Technology for their direct or indirect support and all the lab technicians for their valuable suggestions and for providing excellent opportunities in the completion of this report.

A. G. S. Anirudh	(21341A1201)
B. Jhansi	(21341A1215)
B. Sushma Gayatri	(21341A1213)
Akankhya Gantayat	(21341A1202)

ABSTRACT

Human Activity Recognition (HAR) unlocks the potential to interpret our movements through the power of machine learning. Wearable devices equipped with sensors like accelerometers capture the raw symphony of our actions, translating them into data. This data undergoes a transformation, extracting features like acceleration values to become interpretable by machine learning algorithms. Traditional models like SVMs and even complex deep learning architectures like neural networks are trained on labelled data. These networks, inspired by the human brain, learn the intricate patterns associated with specific activities like walking, running, or standing. By meticulously optimizing models through techniques like hyperparameter tuning, we achieve accurate activity classification. HAR's potential stretches far: fitness trackers can provide personalized coaching based on identified activities, fall detection systems can offer real-time assistance in critical situations, and smart homes can adapt to our movements, creating a personalized and comfortable living environment. As technology continues its relentless march forward, HAR holds immense promise for the future, paving the way for a world where technology seamlessly understands and responds to our every move.

Keywords:

Human activity recognition, Sensor data, Model training and evaluation, Deep learning, Fitness tracking, Activity classification

TABLE OF CONTENTS

LIST OF TABLES

Table of Contents

LIST OF ABBREVIATIONS

ANN	: Artificial Neural Networks
AUC	: Area Under curve
CNN	: Convolutional Neural Networks
CAM	: Computer aided manufacturing
GCN	: Graphical Convolutional Networks
HAR	: Human Activity Recognition
LSTM	: Long Short Tem Memory
LRCN	: Long Term Recurrent Convolutional Network
RNN	: Recurrent Neural Network
UCF50	: University of Central Florida
UCI	: University of California Irvine

1.INTRODUCTION

Human activity recognition (HAR) stands at the forefront of interdisciplinary research, blending elements of computer science, engineering, and behavioral studies to automate the identification and classification of human activities through sensor data analysis. As wearable technology, smartphones, and IoT devices equipped with sensors become increasingly ubiquitous, HAR has emerged as a pivotal fieldwith far-reaching applications across diverse domains such as healthcare, sports analytics, security, and personalized assistance systems. This comprehensive exploration navigates through the foundational concepts, methodologies, challenges, and recent advancements in human activity recognition, offering an in-depth understanding of this rapidly evolving field. Human activity recognition (HAR) encompasses the process of automatically identifying and interpreting human actions and behaviors using computational techniques. It aims to discern, comprehend, and predict human activities based on data collected from a variety of sensors, including accelerometers, gyroscopes, magnetometers, and environmental sensors. By analyzing the patterns and characteristics of sensor data, HAR systems can infer the activities being performed by individuals in real-time or retrospectively.

The significance of human activity recognition spans numerous domains, each with its unique set of challenges and opportunities. In healthcare, HAR facilitates remote patient monitoring, fall detection, and activity tracking for the elderly, promoting independent living and early intervention. Sports analytics leverage HAR to gain insights into athletes' performance, injury prevention, and training optimization. Security systems benefit from HAR by detecting suspicious activities and anomalies in surveillance footage, enhancing situational awareness and threat detection capabilities. Moreover, personalized assistance systems utilize HAR to adapt to users' behaviors and preferences, providing tailored recommendations and support.

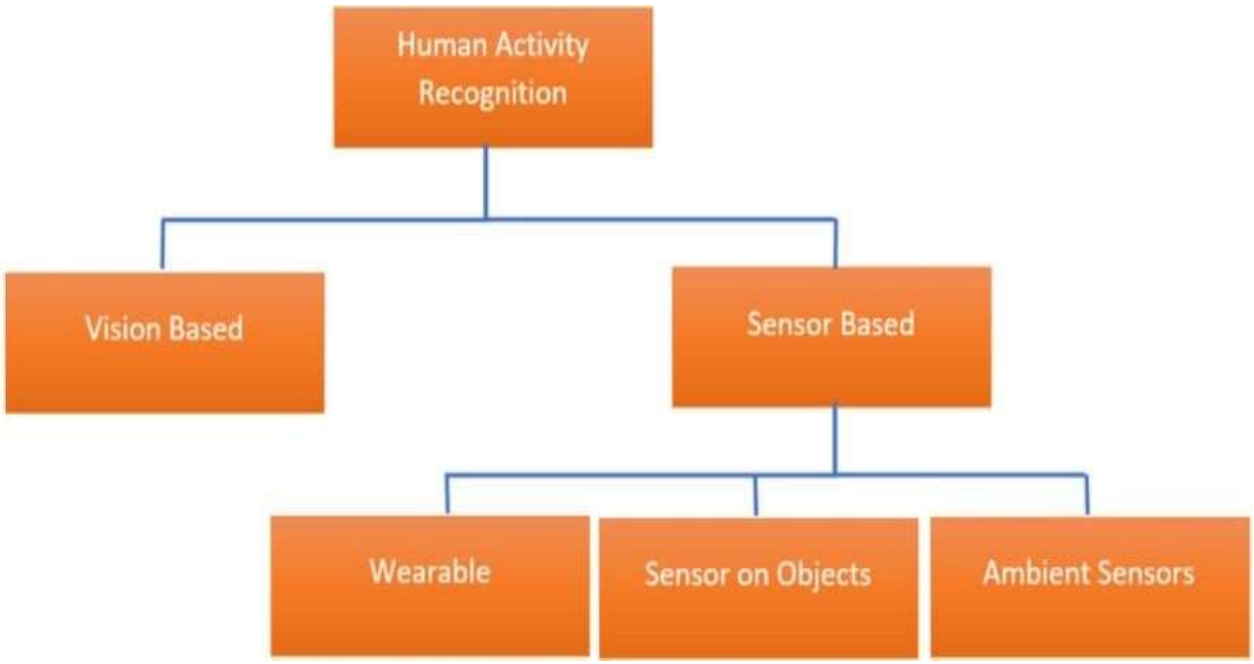


Figure 1.1: HAR Classification

The roots of HAR can be traced back to early research in psychology and neuroscience, where scientists began studying human movement and behavior using observational techniques. However, the advent of wearable technology and the proliferation of smartphones marked a significant inflection point in HAR research. The availability of affordable sensors and advancements in machine learning algorithms fueled rapid progress in HAR, enabling the development of robust activity recognition systems capable of operatingin real-world environments.Various types of sensors play a crucial role in HAR systems, capturing different aspects of human motion and behavior. Accelerometers measure acceleration, gyroscopes detect orientationchanges, and magnetometers sense changes in magnetic fields. Inertial Measurement Units (IMUs) combine multiple sensors to provide comprehensive motion data, while environmental sensors such as cameras and microphones capture contextual information that aids in activity recognition.Data acquisition is a fundamental component of HAR systems, involving the collection of sensor data from individuals performing various activities. Sensors may be placed on the human body or integrated into wearable devices, smartphones, and IoT devices. The collected data undergoes preprocessing, including filtering, noise reduction, and feature extraction, to enhance its quality and relevance for subsequent analysis.Feature extraction transforms raw sensor data into meaningful representations that capture relevant information about human activities. Time-domain features, such as mean, standard deviation, and skewness, provide insights into the statistical properties of sensor signals.

Feature selection aims to identify the most discriminative and informative features for activity recognition while reducing dimensionality and computational complexity. Various techniques, including wrapper methods, filter methods, and embedded methods, are employed to select optimal feature subsets that maximize classification accuracy and generalization performance.

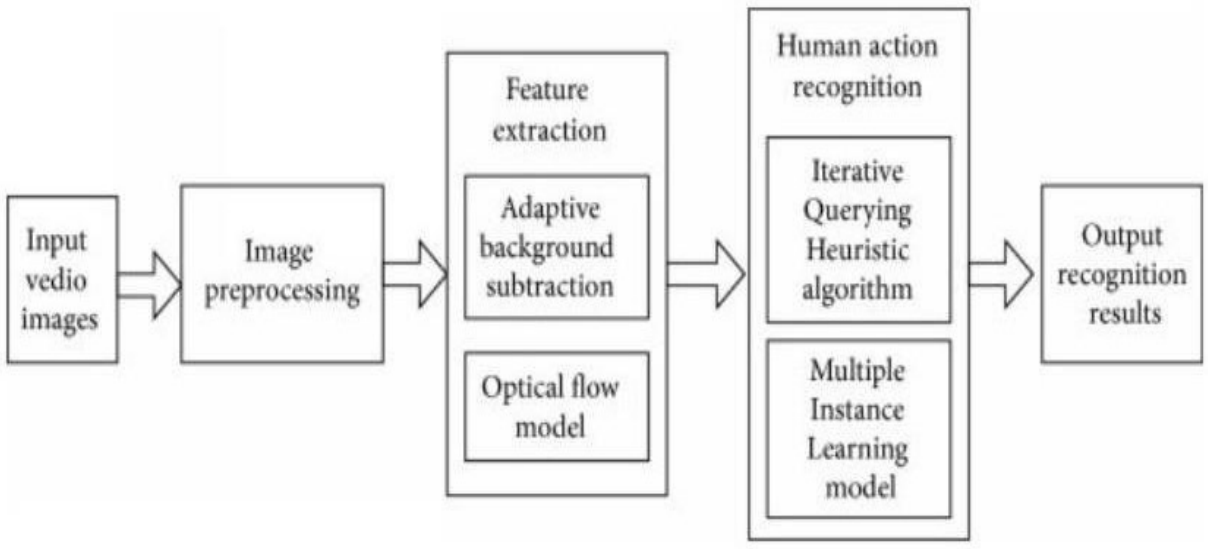


Figure 1.2: Steps Involved in HAR

Supervised learning algorithms, such as Support Vector Machines (SVM), Random Forests, and Artificial Neural Networks (ANN), are widely utilized for activity recognition tasks. These algorithms learn from labeled training data, where each instance is associated with a specific activity label, to classify unseen instances into predefined activity categories. Supervised learning approaches require annotated datasets for training, which may pose challenges in terms of data labelling and scalability. Unsupervised learning techniques, including clustering and anomaly detection, are employed in scenarios where labeled training data is scarce or unavailable. Clustering algorithms group similar instances together based on their feature similarity, enabling the discovery of underlying patterns and structures in unlabeled data. Anomaly detection algorithms identify deviations from normal behavior, thereby detecting unusual or anomalous activities without the need for explicit labeling.

Deep learning models, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have revolutionized activity recognition by automatically learning hierarchical features from raw sensor data. CNNs excel at extracting spatial features from sensor data, while RNNs are well-suited for modeling temporal dependencies and sequential patterns in activity sequences. Deep learning-based HAR systems have achieved state-of-the-art performance across various activity recognition tasks, surpassing traditional machine learning approaches in terms of accuracy and robustness.

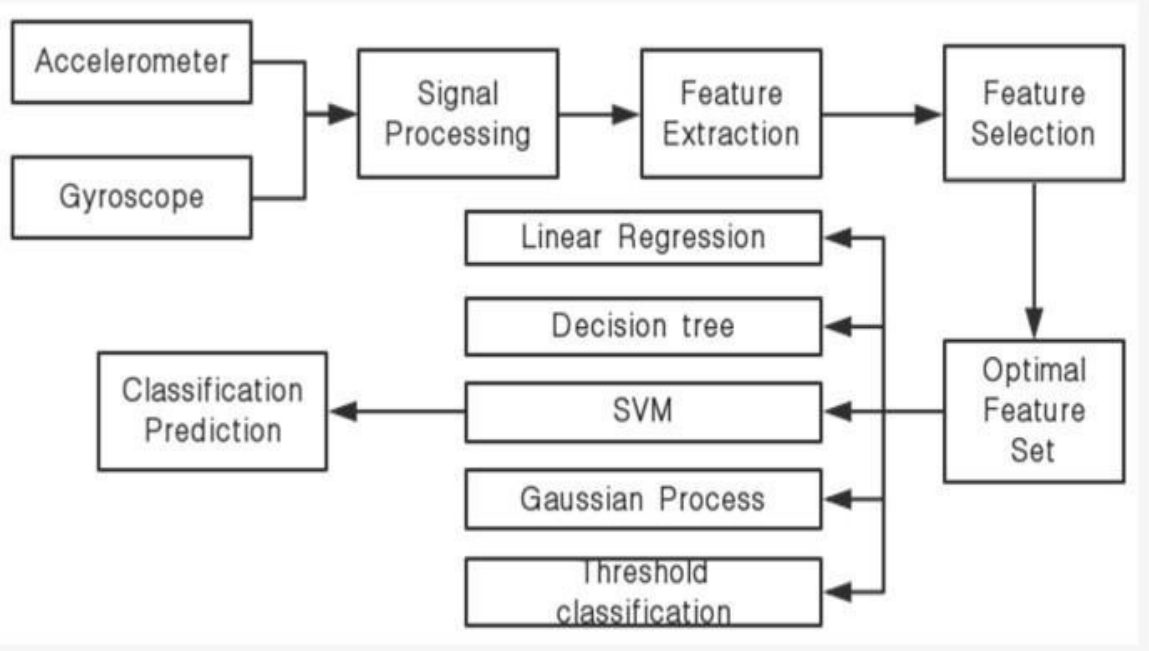


Figure 1.3: A ML System for HAR

Despite significant progress, HAR still faces several challenges that hinder its widespread adoption and deployment in real-world scenarios. These challenges include sensor placement and calibration, data variability and heterogeneity, scalability and adaptability to diverse environments, privacy and security concerns associated with data collection and storage, and interpretability and transparency of machine learning models. The future of human activity recognition holds immense promise, with emerging technologies and research directions poised to address existing challenges and unlock new opportunities. Advancements in sensor technology, including miniaturization, energy efficiency, and multi-modal sensing capabilities, will enable the development of more wearable and unobtrusive HAR systems. Integration of HAR with augmented reality (AR) and virtual reality (VR) technologies will open up novel applications in gaming, training, and immersive experiences.

Furthermore, interdisciplinary collaborations between researchers from computer science, engineering, psychology, and healthcare will foster innovation and driveprogress in HAR research.

Figure 1.4: HAR techniques

Human activity recognition is a dynamic and interdisciplinary field that continues to evolve rapidly, fueled by advances in sensor technology, machine learning algorithms, and application domains. By leveraging data-driven approaches and interdisciplinary collaborations, HAR holds the potential to revolutionize healthcare, sports analytics, security, and personalized assistance systems, ultimately enhancing human well-being and quality of life. This comprehensive exploration serves as a roadmap for researchers, practitioners, and policymakers interested in understanding and contributing to the advancement of human activity recognition.

2. LITERATURE SURVEY

Following research papers are studied in detail to understands the proposed recommendation technique and experimental result for predicting the output.

2.1 Ahmad, W., Kazmi, B. M., & Ali, H. (2019, December). Human activity recognition using multi-head CNN followed by LSTM. In 2019 15th international conference on emerging technologies (ICET) (pp. 1-6). IEEE.

Architecture:

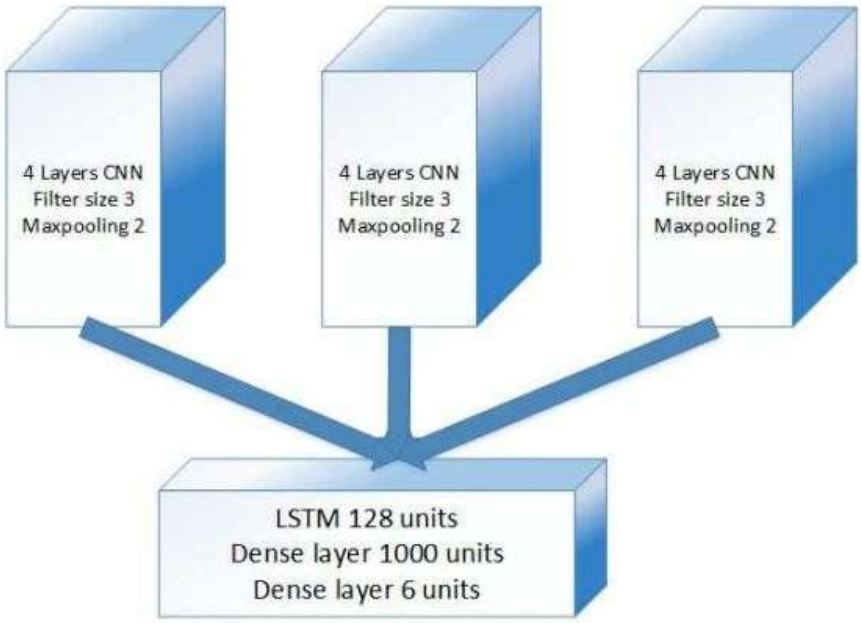


Figure 2.1: Multi-Head CNN-LSTM Architecture

Techniques Used:

The proposed model is trained in an end-to-end method i.e. the parallel CNN along with the LSTM is trained at the same time. The model is tested for 50 epochs, 30 epochs, and 20 epochs.We trained a multihead CNN architecture by using three CNN architectures connected in parallel, Three CNN architectures connected in parallel, as shown in Fig. 4. In which the Single one dimensional CNN discussed above is used in parallel. And the output of all parallel CNNs is merged and provided as input to the LSTM layer.

Then we used the Support Vector Machine, Naive Bayes, Logistic Regression and Random Forest for Sentiment classification.

Data Set Description:

The dataset used for the proposed architecture in this paper is "UCI human activity recognition using smartphone dataset". This dataset was built from the data acquisition from 30 people with age in the range of 19 to 48 years. The individuals were doing daily life work carrying the smartphone at waist position and the smartphone was embedded with the inertial sensors.

Limitations:

There are six actions performed by each person i.e. "Walking, Walking Upstairs, Walking Downstairs, Sitting, Standing, Laying."

Comparison:

Table 2.1: Performance Comparission.

Algorithm Names	Precision	Recall	F1-score	Support
Polynomial SVM	0.84	0.75	0.73	2947
KNN	0.90	0.89	0.89	2947
Single CNN-LSTM	0.95	0.95	0.95	2947
Multi-head CNN-LSTM	**0.96**	**0.96**	**0.96**	**2947**

Table 2.2: Accuracy Scores.

Algorithm	Accuracy
Single CNN-LSTM	94.1 %
Multi-head CNN-LSTM	**95.76 %**

2.2 Zhang P, Zhang Z, Chao H-C. A Stacked Human Activity Recognition Model Based on Parallel Recurrent Network and Time Series Evidence Theory. Sensors. 2020.

Architecture:

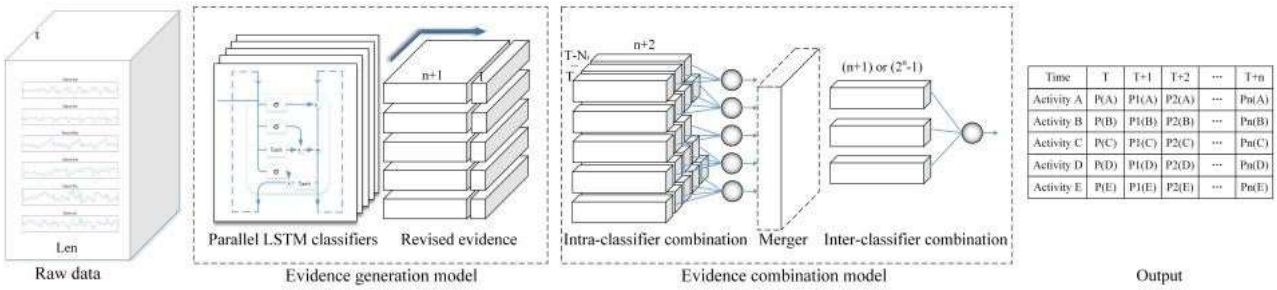


Figure 2.2: Stacked HAR Model.

Techniques Used:

We use a two-layer LSTM network of fixed number of nodes to analyzes the accuracy changing with the window length on UCI-HAR dataset [50], and there is a 50% overlapping between adjacent two windows.

Data Set Description:

The main data set is the UCI-HAR dataset [50] which contains 6 activities of 30 volunteers. Those data were acquired by fixing a smart phone on the waist and 3 channels accelerometer sensor data and 3 channels gyroscope sensor data were recorded simultaneously with a frequency of 50 Hz. The six activities were walking (activity 1), walking up-stairs (activity 2), walking down-stairs (activity 3), sitting (activity 4), standing (activity 5) and lying (activity 6) respectively. This study observe that well trained supervised machine learning technique were able to performvery useful classification on SA polarities.

Limitations:

When it comes to time series data that different categories vary in duration, the window size of input represents the quantity of information, which will be insufficient when the size is too small and will be redundant or incoordinate when the size is too large.

2.3 Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu, and Yunhao Liu. 2018. Deep Learning for Sensorbased Human Activity Recognition: Overview, Challenges and Opportunities. J. ACM 37, 4, Article 111 (August 2018), 40 pages.

Architecture:

The paper introduces a perturbation method that respects human biomechanical constraints to ensure realistic variations in human movement.

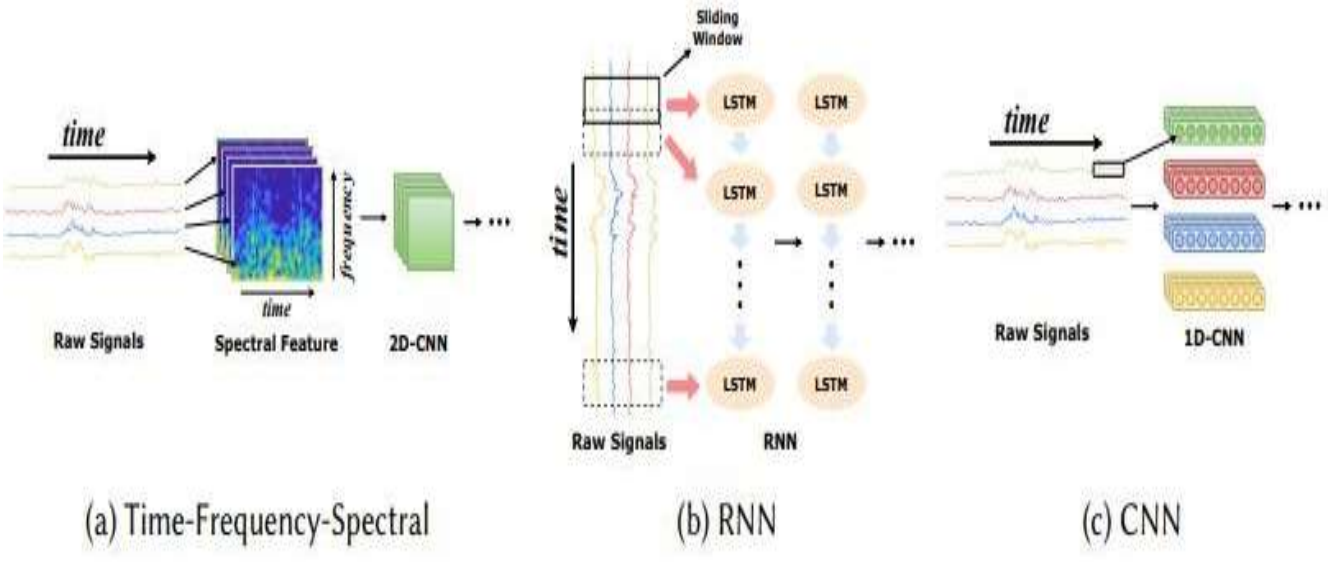


Figure 2.3: Examples of temporal feature extraction models.

Techniques Used:

Figure 2.4: HAR Techniques

Data Set Used:

Table 2.3: DataSet

Dataset	Context	# Subject	# Activities	Sensor Types
WISDM Activity Prediction [75]	Daily Living	29	6	Wearable
UCI HAR [8]	Daily Living	30	6	Wearable
OPPORTUNITY [26, 126]	Daily Living	4	9	Wearable, Object, Ambient

2.4 Human Activity Recognition With Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review E. Ramanujam , Thinagaran Perumal , Senior Member, IEEE, and S. Padmavathi(2022) **Architecture:**

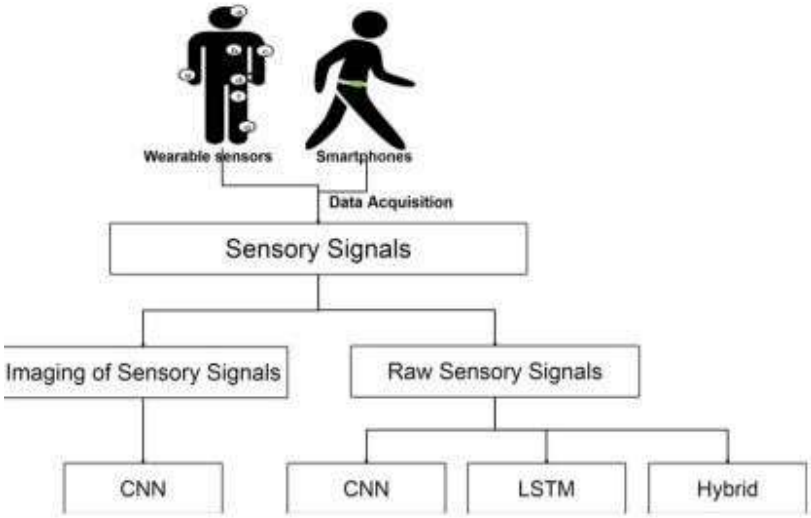


Figure 2.5: Categorization of proposed DL models

Techniques Used: CNN, LSTM, Hybrid Models

Data Set Description:

Table 2.4: DataSet Description

Dataset name	Sensor (s) data	# of Activities		Type of Activity			# of Subjects	Sampling Frequency	Device(s) used	Device Position	Environment
		ADLs	Falls	Simple	Complex	Postural					
PAMAP2 [12]	T, A, G, O	12	×	✓	✓	×	9	100Hz	3 IMU units 1 Heart rate monitor	Wrist, chest & dominant side's ankle	Controlled Lab
HHAR [13]	A, G	6	×	✓	×	×	9	Highest	4 Smartwatches & 8 Smartphone	Smartphone - Waist, Pouch	Out of Lab
MHEALTH [14-15]	A, G, M, ECG signals	12	×	✓	×	×	10	50Hz	Shimmer2 wearable sensors	right wrist, left ankle, and chest	Out of Lab
UCI-HAR [16]	A, G	6	×	✓	×	×	30	50Hz	Smartphone	Left belt and No specific position	Controlled

Limitations:

- Complex activities and Postural Transitions.
- Cross-Adaptability data.
- limited contextual data for training.
- Hyperparameter optimizations.

2.5 From Movements To Metrics: Evaluating Explainable Ai Methods In SkeletonBased Human Activity Recognition Kimji N. Pellano1 , Inga Strumke ¨ 2 , and Espen Alexander F. Ihlen1 (2023) **Architecture:**

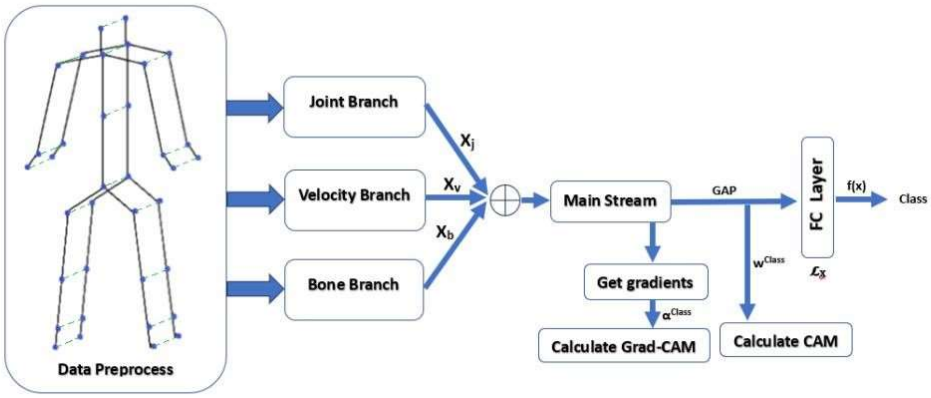


Figure 2.6 : The efficient GCN pipeline showing the variables for calculating faithfulness and stability. pertubation is performed in data preprocess stage.

Techniques Used:

Class Activation Mapping (CAM) was used in EfficientGCN [7] and ST-GCN [8] to highlight the body points significant for specific actions. In [9], Gradient-weighted Class Activation Mapping (Grad-CAM) was implemented in ST-GCN. There is a growing trend towards using explainable AI (XAI) methods, extending from CNNs to ST-GCNs, yet XAI metrics to assess their reliability in this domain have yet to be tested.

Data Set Description:

The NTU RGB+D-60 dateset contains 60 action classes with over 56 thousand 3D skeleton data, each composed of sequential frames captured from 40 different subjects using the Kinect v2 camera with depth sensor.

Limitations:

- It did not offer insights into its performance relative to other XAI methods. Moreover, their choice of using masking/occlusion to check for changes in prediction output raises concerns.
- Lack of detailed analysis on user engagement and long-term impact

3.METHODOLGY

The full workflow of our human activity recognition model is depicted in Figure X. First, we gathered video data from a public source. Before partitioning the video dataset into training and testing data, various preprocessing techniques such as filtering, shrinking frames, and normalizing data are applied. On the processed data, we trained numerous machine learning models, including CNN, LSTM, ConvLSTM, and LRCN, where state-of-the-art works did not use LRCN. Finally, we conducted a thorough performance analysis. For activity recognition, our model may upload testing video data from YouTube and other sources. The experiments reveal that LRCN outperformed all other models except CNN in terms of accuracy.We describe our approach step-by-step in the following paragraphs.

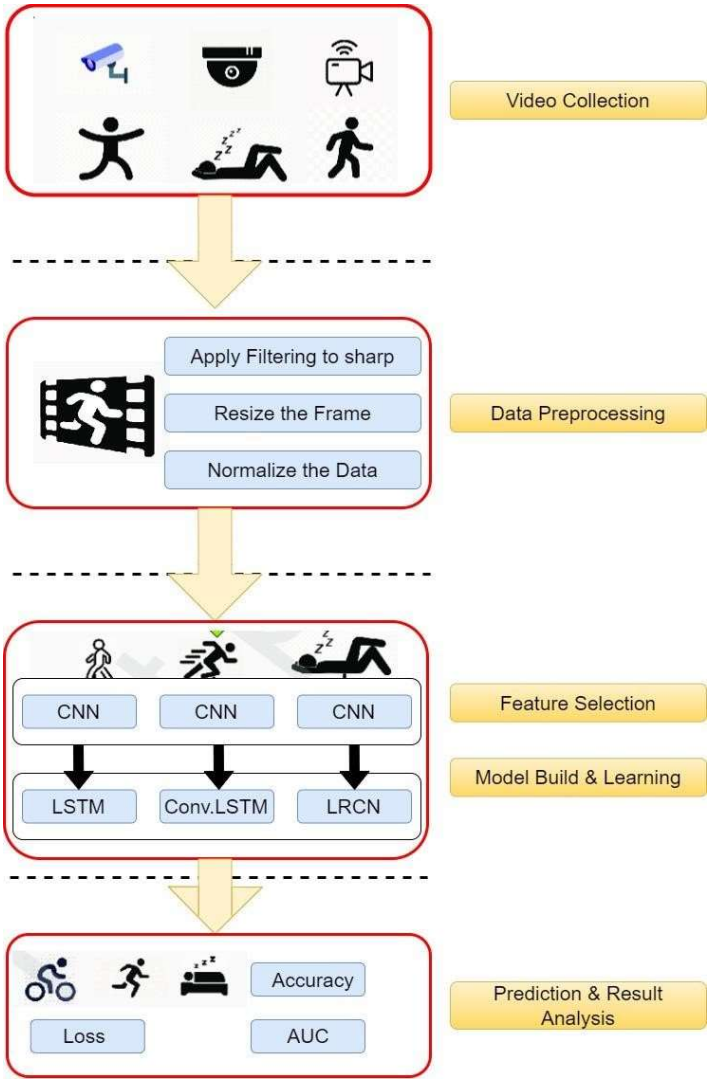


Figure 3.1: Proposed Model Architecture

- **Input Dataset**: We have used UCF-50 to train our model. We need photos or videos to detect human activity with ML or DL models. The activities in the datasets include strolling, running, leaping, playing tennis, walking with a dog, etc. Such data can be captured by an on-board computer system, external camera, or a variety of surveillance cameras.

- **Pre-processing**: The Dataset Preprocessing is essential before feeding the datasets to the learning module as real-world data is generated with noise and unwanted substances. To obtain better outcome from the data, some preprocessing on the dataset is performed in our model. We first read the video streams from the dataset, and then we reformat the video frames to a predetermined length and width to accelerate convergence when training the network.

- **Classification of Human Activity**: The following step are performed in order to construct the human activity recognition model.

❖ The Video clips datasets are collected to train the human activity

❖ The processing are performed on videos applying filter to sharp and resize the frames to a given width and height, thus normalizing the data.

❖ The normalized data are split into 80%-20% for training and testing purpose. 20% of the videos in our dataset are used for validation.

❖ After that four Deep Learning models such as CNN, LSTM, Conv LSTM and LRCN are trained to learn human activity.3.1. Deep Learning models We discussed four Deep Learning Models such as Convolutional Neural Network, Long Short Term Memory, Convolutional Long Short Term Memory, and Long Term Recurrent Convolutional Network.

- **Convolutional Neural Network (CNN)**: Deep convolutional neural networks are used to evaluate visual imagery. Convolution is a mathematical operation that, given two functions, generates a third function that demonstrates how the shape of one function is altered by the other. Instead of having a fully linked layer foreach pixel, CNNs only have enough weights to examine small image portions at a time. In the majority of the cases, a convolution layer precedes a pooling and activation function. A Convolutional Neural Network consists of many layers. A CNN has three layers: a convolutional layer, a pooling layer, and a fully connected layer.

- **CNN + LSTM Approach:** An LSTM network is especially built to operate with a data sequence when creating an output since it checks all previous inputs. Although LSTMs are a Recurrent Neural (RNN) variant, RNNs are not known to be successful in dealing with lengthy dependencies

in input sequences owing to an issue known as the Disappear gradients difficulty. For prolonged input sequences, LSTMs were expected to solve the vanishing gradient and remember frame of reference. This improves a Classifier's capacity to tackle issues with sequential data, such as time series prediction, voice recognition, language translation, and music creation. LSTM produces better outcome for human activity detection from videos. A CNN extracts spatial attributes at a certain time step in the input pattern (videos), followed by an LSTM to detect temporal correlations between frames. Therefore, we train two combinations of CNN and LSTM:LRCN and ConvLSTM to accomplish Action Recognition while using the Spatial-Temporal aspect of the clips as shown in Figure X.

- **Long-Term Recurrent Convolutional Network (LRCN) Approach:** In 2016 a group of authors suggested end-to-end trainable class of architectures for visual recognition and description. The main idea is to use a combination of CNNs to learn visual features from video frames and LSTMs to transform a sequence of image embeddings into a class label, sentence, probabilities, or whatever you need. Thus, raw visual input is processed with a CNN, whose outputs are fed into a stack of recurrent sequence models. As it is described in the figure below, LRCN processes the variable-length visual input with a CNN. And their outputs are fed into a stack of recurrent sequence models which is LSTM in the figure. The final output from the sequence models is a variable-length prediction. This makes LRCN is proper models to handle tasks with time-varying inputs and output, such as activity recognition, image captioning and video description. Below figure is task-specific instantiations of LRCN model for each task.

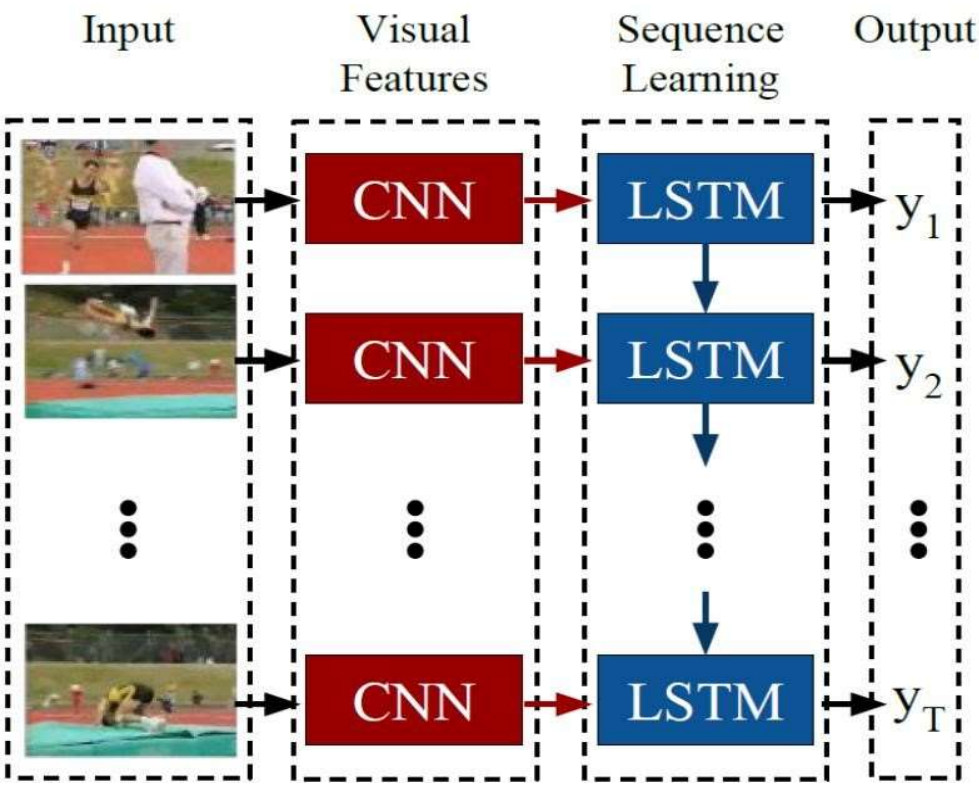


Figure 3.2 : Proposed Model Architecture

The LRCN Approach is implemented in an unified framework by combining Convolution and LSTM layers.The use of a CNN model with an LSTM model trained separately is also a viable option. It is possible to utilize a pre-trained model to extract spatial data from video frames using CNN, and it is possible to delicate this model for the application. As a result, the LSTM prototype may utilize the information collected from the video by CNN to make predictions about the activity being performed in the video. The Long-term Recurrent Convolutional Network (LRCN), on the other hand, incorporates CNN and LSTM layers into one model. At each time step, the collected spatial features are given to an LSTM layer for temporal sequence modeling, which begins with the Convolutional layers. A robust model is produced as a result of the network learning spatiotemporal properties in this way during an endto-end training session.

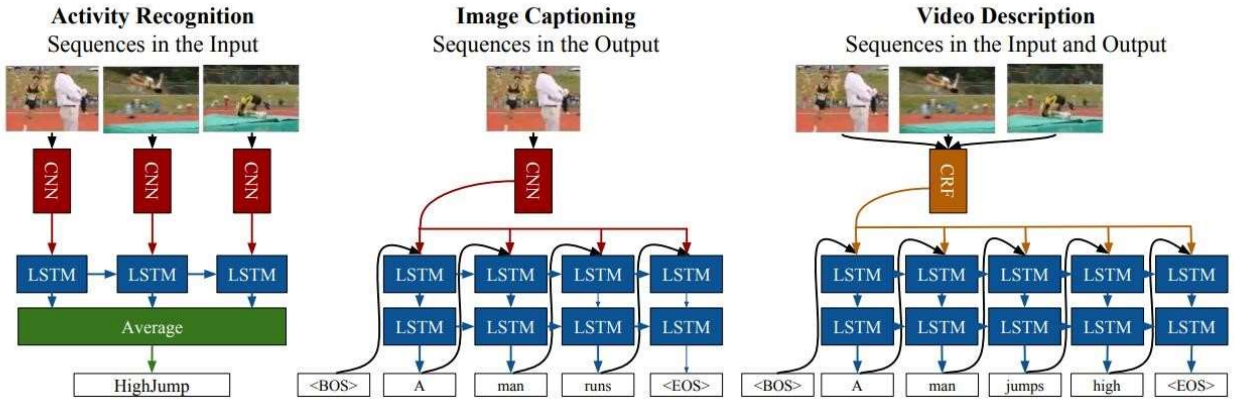


Figure 3.3: LRCN Approach

LSTM Layer :

The central role of an LSTM model is held by a memory cell known as a 'cell state' that maintains its state over time. The cell state is the horizontal line that runs through the top of the below diagram. It can be visualized as a conveyor belt through which information just flows, unchanged.Information can be added to or removed from the cell state in LSTM and is regulated by gates. These gates optionally let the information flow in and out of the cell. It contains a pointwise multiplication operation and a sigmoid neural net layer that assist the mechanism. The sigmoid layer gives out numbers between zero and one, where zero means 'nothing should be let through', and one means 'everything should be let through'.

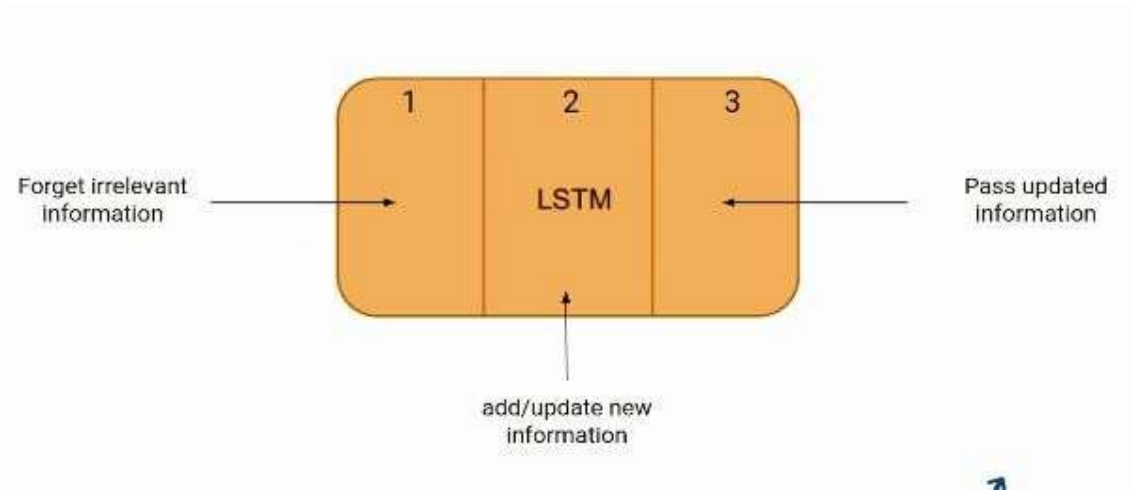


Figure 3.4: LSTM Architecture

- **Convolutional Long Short-Term Memory (ConvLSTM) Approach:**

In this paper, we use ConvLSTMcells to implement the model. Cells in a ConvLSTM network have convolution operations built into the network. It's an LSTM with built-in convolution, so it can detect spatial aspects in the data while also taking the temporal relationship into consideration. To classify video, this technique effectively captures both the spatial and temporal relationships between the frames. Keras ConvLSTM2D recurrent layers are used to build the model. The number of filters and kernel size necessary to perform convolutional operations are also taken into account by the ConvLSTM2D layer in this scheme. Softmax activation is used in the Dense layer, which processes each action category's chance of occurring.

- **Long-Term Recurrent Convolutional Network (LRCN) :**

TheThe LRCN Approach is implemented in an unified framework by combining Convolution and LSTM layers.The use of a CNN model with an LSTM model trained separately is also a viable option. It is possible toutilize a pre-trained model to extract spatial data from video frames using CNN, and it is possible to delicatethis model for the application. As a result, the LSTM prototype may utilize the information collected from the video by CNN to make predictions about the activity being performed in the video. The Long-term Recurrent Convolutional Network (LRCN), on the other hand, incorporates CNN and LSTM layers into one model. At each time step, the collected spatial features are given to an LSTM layer for temporal sequencemodeling, which begins with the Convolutional layers. A robust model is produced as a result of the networklearning spatiotemporal properties in this way during an end-to-end training session. An LRCN model is

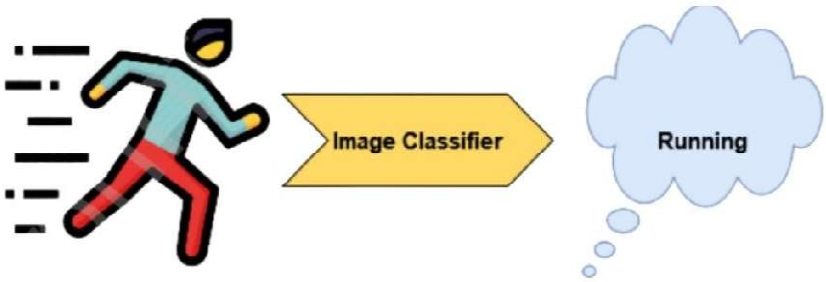


Figure 3.5: Dataset Clips Example

At each time step, the collected spatial features are given to an LSTM layer for temporal sequence modeling, which begins with the Convolutional layers. A robust model is produced as a result of the network learning spatiotemporal properties in this way during an end-to-end training session. An LRCN model is shown below.

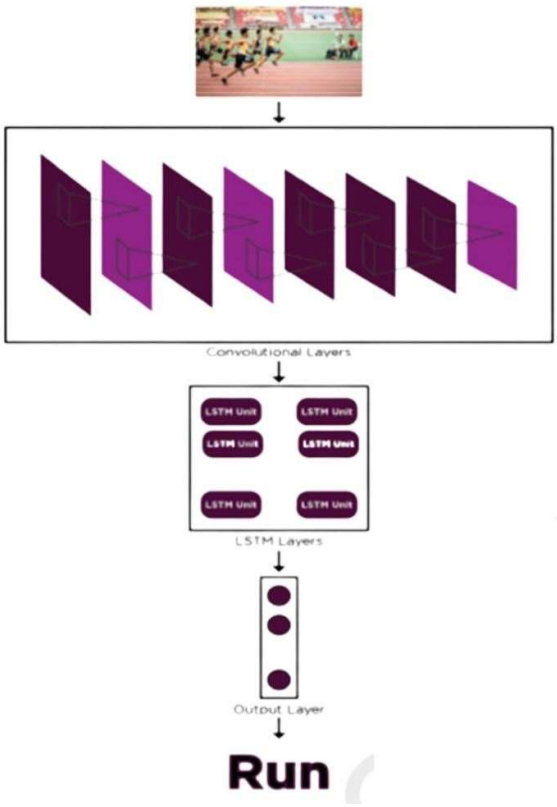


Figure 3.6: LRCN Approach

LRCN is a class of architectures which combines Convolutional layers and Long Short-Term Memory (LSTM).

BASIC LRCN

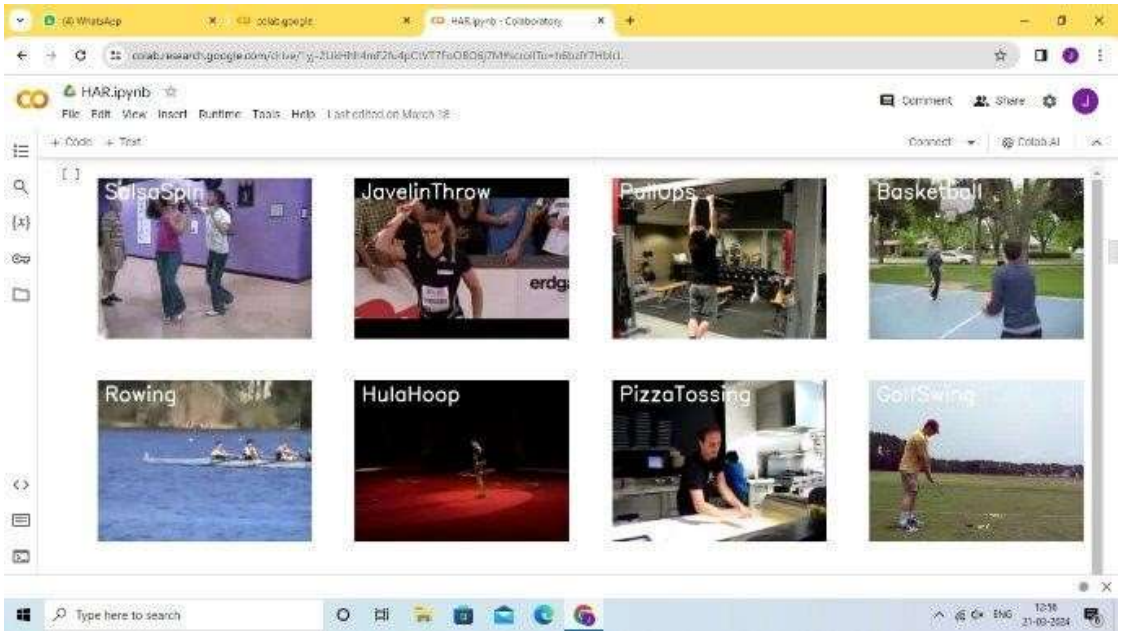
- Convolutional2D Layer
- LSTM Layer
- Dense Layer [fully - connected]

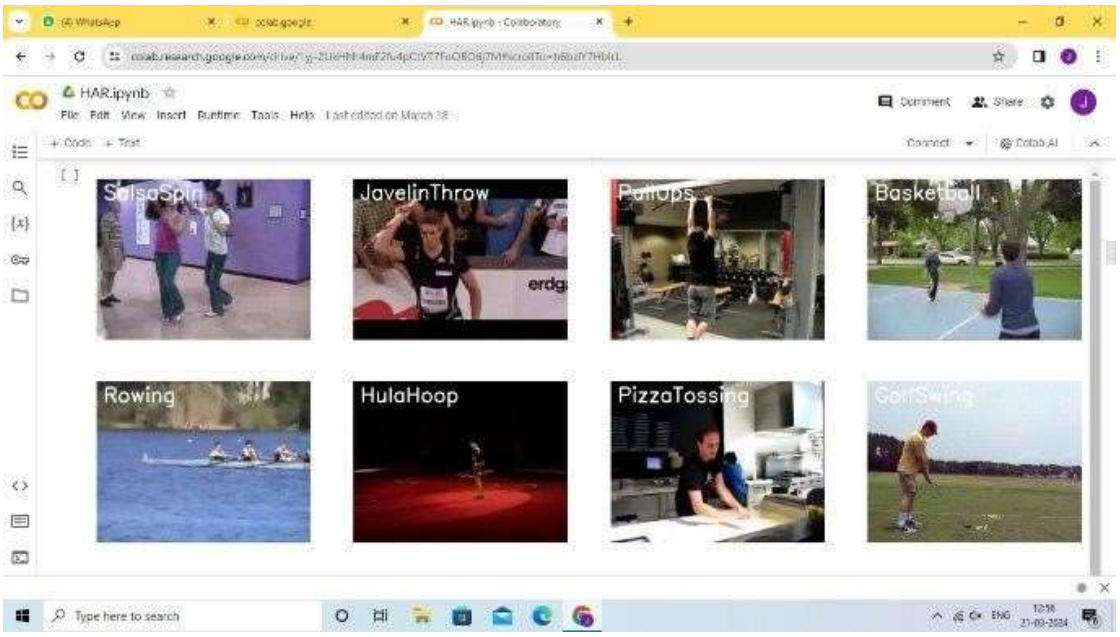
ADVANCED LRCN

- 3 Convolutional2D Layers
- LSTM Layer
- Dense Layer [fully - connected]

Data Set :

UCF50 is an action recognition data set of realistic action videos, collected from YouTube, having 50 action categories. This data set is an extension of YouTube Action data set (UCF11) which has 11 action categories. Most of the available action recognition data sets are not realistic and are staged by actors. In our data set, the primary focus is to provide the computer vision community with an action recognition data set consisting of realistic videos which are taken from YouTube. Our data set is very challenging due to large variations in camera motion, object appearance and pose, object scale, viewpoint, cluttered background, illumination conditions, etc. For all the 50 categories, the videos are grouped into 25 groups, where each group consists of more than 4 action clips. The video clips in the same group may share some common features, such as the same person, similar background, similar viewpoint, and so on. UCF50 data set's 50 action categories collected from YouTube are: Baseball Pitch, Basketball Shooting,
Bench Press, Biking, Biking, Billiards Shot, Breaststroke, Clean and Jerk, Diving, Drumming, Fencing, Golf Swing, Playing Guitar, High Jump, Horse Race, Horse Riding, Hula Hoop, Javelin Throw, Juggling Balls, Jump Rope, Jumping Jack, Kayaking, Lunges, Military Parade, Mixing Batter, Nun chucks, Playing Piano, Pizza Tossing, Pole Vault, Pommel Horse, Pull Ups, Punch, Push Ups, Rock Climbing Indoor, Rope Climbing, Rowing, Salsa Spins, Skate Boarding, Skiing, Ski jet, Soccer Juggling, Swing, Playing Tabla, TaiChi, Tennis Swing, Trampoline Jumping, Playing Violin, Volleyball Spiking, Walking with a dog, and Yo Yo





.

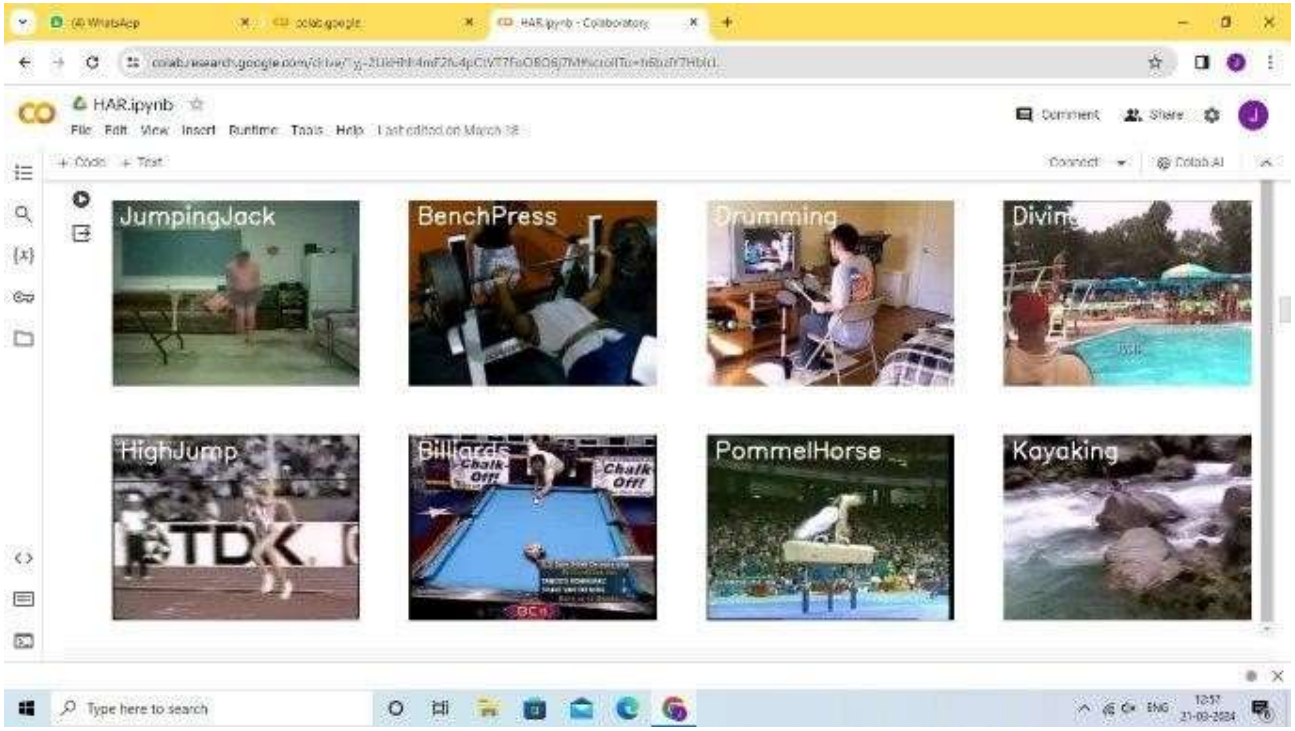


Figure 3.7: UCF 50 Dataset Clips Example

4.RESULTS

Accuracy:

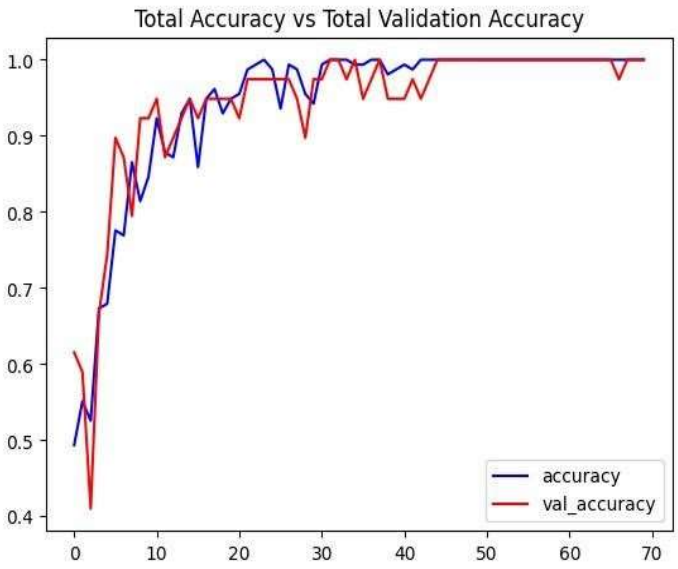


Figure 4.1: Accuracy

The graph showing the accuracy of an LRCN model, it compares two types of accuracy: total accuracy and total validation accuracy.

Total accuracy:

Total accuracy represents how well the LRCN model performed on the training data.

Total validation accuracy:

Total validation accuracy represents how well the LRCN model performed on a separate set of data, the validation data, that the model was not trained on.

The graph shows both total accuracy (red line) and total validation accuracy (blue line) on the yaxis, with the x-axis representing the number of training epochs.The part of the x-axis indicating the number of epochs is cut off in the image, making it difficult to see the complete picture.

General observations:

Both the total accuracy and validation accuracy appear to be increasing, which is a positive sign. The total accuracy is consistently higher than the validation accuracy. This could be a sign of overfitting, where the model is performing well on the training data (high total accuracy) but may not perform as well on unseen data (represented by the validation accuracy).As the epochs are increasing the acuuraccy of the model is decreasing.

The reasons why overfitting is happening:

- The LRCN model being too complex for the amount of training data available.
- The training data may not be representative of real-world data

Loss:

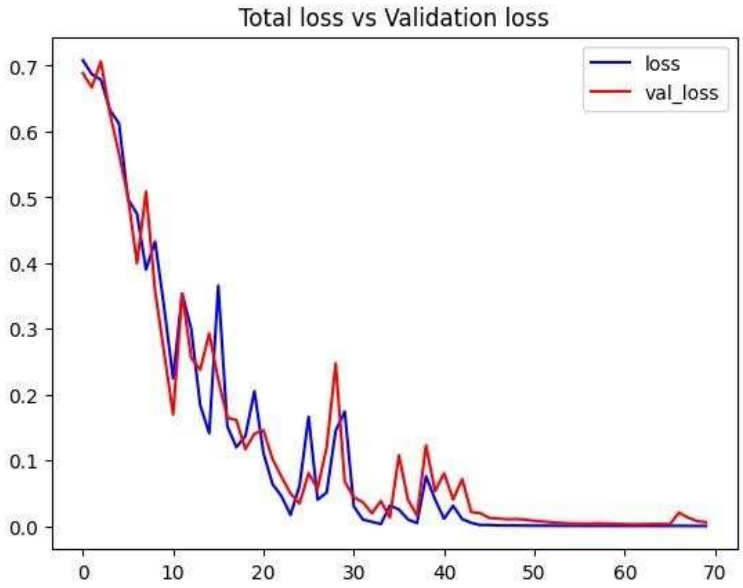


Figure 4.2: Dataset Clips Example

The graph showing the loss of an LRCN model, it compares two types of accuracy: total loss and total validation loss.

Total loss:

Total loss represents the performance of the model on the training data.

Validation loss:

validation loss is a metric used to assess how well a model performs on unseen data. It helps prevent over fitting, a situation where the model performs very well on the training data but poorly on new data.

General observations:

The total loss is consistently higher than the validation loss. This suggests that the model may be over fitting the training data. Over fitting is when a model performs well on the training data but poorly on unseen data.As the epochs are increasing the loss of the model is increasing .

Metrics:-

The LRCN model is performing binary classification (e.g., classifying videos as containing a specific action or not), so AUC (Area Under the Curve) is used as an evaluation metric.AUC would assess the model's ability to distinguish between positive videos (containing the action) and negative videos (not containing the action).

How AUC would be used:

Classification Scores: The LRCN model would likely output a score for each video it processes. This score can represent the probability of the video belonging to the positive class (violence in this case).

Thresholding: A classification threshold is chosen. Videos with scores above the threshold are classified as positive (containing violence), while those below are classified as negative (not containing violence).

ROC Curve: By varying the classification threshold, we can calculate the True Positive Rate (TPR) and False Positive Rate (FPR) for each threshold. Plotting TPR on the y-axis and FPR on the x-axis creates the ROC Curve.

AUC Calculation: The AUC metric then calculates the area under the ROC Curve. A higher AUC value (closer to 1) indicates better performance, meaning the model can effectively differentiate between positive and negative video classifications.

Recognition:

Figure 4.3: Detecting Activity

5.FUTURE SCOPE

To develop full potential of deep learning in human activity recognition, some future research directions are worthy of further investigation. Future directions can be stimulated by the challenges summarized in this work. Despite the effort devoted to these challenges, some of them are still not fully explored such as class imbalance, composite activities, concurrent activities, etc. Although current research works still lack comprehensive and reliable solutions for the challenges, they lay concrete foundations and show guidance for future directions. Moreover, there are other research directions that have rarely been explored before. We outline several key research directions that urgently need to be exploited as follows.

- **Independent unsupervised methods :** Human activity recognition needs a sufficient amount of annotated samples to train the deep learning models. Unsupervised learning can help mitigate such requirements. So far, deep unsupervised models used for human activity recognition are mainly used for extracting features but are not able to identify activities because there is no ground truth. Therefore, one potential method for unsupervised learning to infer true labels is to seek other knowledge, which leads us to a popular method, deep unsupervised transfer learning [18]. Another way is to resort to data-driven methods such as ontology .

- **Identifying new activities:** Identifying novel activities that have never been seen by the models is a big challenge in human activity recognition. A reliable model should be able to learn the new knowledge online and achieve accurate recognition without any ground truth. A promising way is to learn features that are scalable to diverse activities. While enlightens us that mid-level attributes can be used to depict activities with a set of characteristics, disentangled features may be another serviceable solution to representing novel activities.

- **Future activity prediction:** Future activity prediction is an extension of activity recognition. Unlike activity recognition, the activity prediction system can forecast users' behaviors in advance. The prediction system is useful in detecting human intention so it can be applied to smart services, criminal detection and driver behavior prediction. In some common behavior tasks, the activities are usually in a certain order. Therefore, modeling the temporal dependencies across activities is beneficial to predict future predictions. LSTMs are suitable for such tasks. But for long-span activities, LSTMs cannot contain such long dependencies. In this case, intention recognition based on brain signals can assist to inspire activity prediction.

6.CONCLUSION

We have used LRCN model to perform Human Activity Recognition.Unleashing the power of human activity recognition in videos, LRCN models have emerged as the reigning champions. These models are a fusion of two deep learning techniques: convolutional neural networks (CNNs) that excel at image recognition and long short-term memory (LSTM) networks that master sequence analysis. This dynamic duo empowers LRCNs to not only identify how someone looks (spatial features) but also decipher their movements (temporal features). This allows them to crack the code of complex activities, like distinguishing between walking and dancing. Training these champions requires a significant investment: a vast library of labeled videos and serious processing power. However, the rewards are substantial. LRCN models offer a powerful and adaptable solution for various human activity recognition tasks in videos. They can be fine-tuned for specific applications, making them valuable tools across diverse fields. As research progresses, we can expect even greater efficiency, robustness, and adaptability from these future superstars of video analysis.bayes. We have taken two datasets,where the first dataset which is having the high length in the sentence took time when compared to the other dataset whose length of the sentence is less. Naive bayes model isconsidered as a time efficient as the model takes the less time for the training the model. Support Vector Machine (SVM) is a memory efficient as the only stores the support vectors data points only. Therefore,SVM uses when the user has less idea on the data.

7.REFERENCES

(1) Ahmad, W., Kazmi, B. M., & Ali, H. (2019, December). Human activity recognition using multihead CNN followed by LSTM. In *2019 15th international conference on emerging technologies (ICET)* (pp. 1-6). IEEE.

(2) Zhang P, Zhang Z, Chao H-C. A Stacked Human Activity Recognition Model Based on Parallel Recurrent Network and Time Series Evidence Theory. *Sensors*. 2020

(3) Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu, and Yunhao Liu. 2018. Deep Learning for Sensor based Human Activity Recognition: Overview, Challenges and Opportunities. J. ACM 37, 4, Article 111 (August 2018)

(4) Ramanujam, E. and Perumal, Thinagaran and Padmavathi, S. (2021) *Human activity recognition with smartphone and wearable sensors using deep learning techniques: a review.* IEEE Sensors Journal, 21 (12). 13029 - 13040. ISSN 1530-437X

(5) Pellano, K.N.; Strümke, I.; Ihlen, E.A.F. From Movements to Metrics: Evaluating Explainable AI Methods in Skeleton-Based Human Activity Recognition. Sensors 2024,24,1940