iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

HUMAN ACTIVITY RECOGNITION USING MACHINE LEARNING

LEKHANA YADAGIRI[1] (MCA student),
M. BALA NAGA BHUSHANAMU[2] (Asst. Prof)
Department of Information Technology & Computer application,
Andhra University College of Engineering, Visakhapatnam, AP.
Corresponding Author: Lekhana Yadagiri
lekhanaml2106@gmail.com

ABSTRACT:
Human activity recognition and classification are among the most critical challenges in the field of computer science. The ability to detect and interpret human actions is fundamental to the development of intelligent video systems. These systems play a significant role in a wide range of applications, including human-computer interaction, security surveillance, and healthcare monitoring.
Despite substantial advancements, accurately identifying human activities in diverse and dynamic environments remains a complex issue. This article explores current research in human activity recognition, focusing on various cognitive and computational approaches.

KEYWORDS:
Computer Version, Intelligence video Systems, Security, Health Care, Vision-Based Methods, Smartphones, Classification, Challenges in HAR, Surveillance, Accuracy Comparison.

I. INTRODUCTION

Skeleton-based features have become increasingly prominent in the fields of human action recognition and human-computer interaction. This approach focuses on identifying and tracking key points of the human body from video or image data, often using depth cameras and specialized sensors. By capturing the motion of joints and limbs, these systems can analyze movement patterns through the application of computer vision and machine learning techniques. Such technology finds broad applications in areas like virtual reality, gaming, healthcare, and security monitoring.
Despite their advantages, recognizing human actions using skeleton data presents notable challenges. These include handling multiple individuals in a scene, diverse camera angles, and complex human interactions. Early systems largely depended on handcrafted features and spatiotemporal modelling. However, recent breakthroughs in deep learning have enabled more accurate recognition by leveraging skeleton key points and advanced spatio-temporal analysis.
One such advancement is the MS-G3D network, which uses 3D convolutional layers and attention mechanisms to automatically extract relevant features from skeletal sequences. In another development, the 3D-GSD method (3D Graph Convolutional Feature Selection and Dense Pre-estimation) introduces both spatial and temporal attention mechanisms. These methods aim to capture both localized and global aspects of movement, improving the analysis of complex human poses and behaviors.
Additionally, research is exploring the effectiveness of Convolutional Neural Networks (CNNs) for recognizing human behavior from various camera angles. In recent years, smartphones have also emerged as accessible platforms for activity recognition, thanks to their built-in sensors. This area has become a key focus for modern research.
This study presents a comprehensive review of existing literature on human activity recognition. A dataset from the UCI Machine Learning Repository was used to develop Artificial Intelligence (AI) models capable of identifying human activities. The dataset, being multivariate, was analysed using several classification algorithms including Random Forest, K-Nearest Neighbors (KNN), Neural Networks, Logistic Regression, Stochastic Gradient Descent (SGD), and Naïve Bayes.
To optimize performance, a feature selection process was applied to reduce dataset dimensionality. Each model's performance was evaluated using Precision, Recall, and Confusion Matrices. The results indicated that Neural Networks and Logistic

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Regression delivered the highest accuracy, though they required more computational power and memory compared to lighter models like KNN or Naïve Bayes.
The entire workflow starts with data acquisition from sources like cameras, wearables, or smartphones. This raw data, often noisy or inconsistent, is pre-processed through filtering, normalization, and segmentation into fixed time frames. Next, feature extraction is performed—either manually (e.g., mean, variance, FFT) or automatically via deep learning models that learn features directly from the raw data.
Finally, classification algorithms are applied. Traditional machine learning models like SVM, Random Forest, and KNN are used when features are pre-engineered. In contrast, deep learning architectures such as CNNs, LSTMs, or hybrid CNN-LSTM models can perform end-to-end learning, processing raw input directly for activity recognition. Real-time implementation may involve integrating these models with cameras or deploying them on embedded systems.

II. RELATED WORK:

2.1 Sensor-Based HAR

Sensor-based approaches are among the earliest and most practical methods for Human Activity Recognition. These use data from accelerometers, gyroscopes, and magnetometers in smartphones or wearable devices. **Anguita et al. (2013)** utilized the UCI HAR dataset with Support Vector Machines (SVM) to classify six common activities.
Kwapisz et al. (2011) used mobile phone accelerometer data and machine learning algorithms like Decision Trees and KNN, proving real-time HAR feasibility on handheld devices.

2.2 Vision-Based HAR

Vision-based HAR leverages video sequences or image frames to identify actions using spatial and temporal patterns. These methods are powerful but resource-intensive.

- Ji et al. (2013) proposed a 3D Convolutional Neural Network (3D-CNN) for learning motion features from video clips.
- Simonyan and Zisserman (2014) developed the Two-Stream CNN to separately process spatial (RGB) and motion (optical flow) inputs for improved recognition**.**

2.3 Deep Learning-Based Models

Deep learning has become a dominant approach in HAR, allowing automatic feature learning from raw data.

- CNNs are effective for extracting spatial features.
- LSTMs model temporal dependencies in sensor or video sequences.
- Hybrid CNN-LSTM models have been successfully applied in HAR to combine the strengths of both architectures for better performance.

2.4 Hybrid and Multimodal Approaches

Hybrid approaches that fuse multiple data types—such as sensor and video input—enhance recognition accuracy and robustness.

- These models are designed to overcome limitations of single-modality systems, making them suitable for complex environments.
- Multimodal deep learning, often using attention mechanisms or ensemble models, is increasingly adopted for real-world HAR systems.

2.5 Review of Related Work

Numerous studies have explored Human Activity Recognition (HAR) using both sensor-based and vision-based approaches. Early research focused on traditional machine learning techniques using data from accelerometers and gyroscopes embedded in smartphones and wearable devices. For example, Anguita et al. applied Support Vector Machines (SVM) on the UCI HAR dataset, achieving significant accuracy in recognizing basic activities like walking and sitting. Similarly, Kwapisz et al. used smartphone accelerometer data to classify activities using Decision Trees and K-Nearest Neighbors, highlighting the feasibility of mobile-based HAR.
With advancements in deep learning, vision-based methods gained popularity. Ji et al. introduced 3D Convolutional Neural Networks (3D-CNNs) to process spatial and temporal video data, while Simonyan and Zisserman proposed a Two-Stream CNN architecture that separately analyzes RGB frames and optical flow for more accurate action recognition. These models improved precision but introduced computational and privacy concerns.
Recent works focus on hybrid models, such as CNN-LSTM combinations, which capture both spatial and temporal patterns in sequential activity data. Multimodal systems that integrate both sensor and vision data are also being developed for greater

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

accuracy in real-world environments. Despite these advances, challenges like inter-subject variability, noisy data, and real-time deployment limitations persist. As a result, current research is moving toward transformer-based architectures, self-supervised learning, and federated models to enhance adaptability, scalability, and privacy in HAR systems.

2.6 Contribution of Current Work

The current work contributes to the field of Human Activity Recognition (HAR) by proposing a hybrid approach that leverages both sensor-based and deep learning techniques for accurate and real-time activity classification. Unlike traditional HAR systems that rely solely on handcrafted features or single-model architectures, this project integrates **CNN-LSTM deep learning models** to capture both spatial and temporal patterns in activity data.

Key contributions include:

- **Hybrid Deep Learning Model**: Implementation of a CNN-LSTM architecture that extracts spatial features from raw sensor data using convolutional layers and models temporal dynamics through LSTM units, leading to improved recognition accuracy for sequential activities.
- **Multi-Source Data Handling**: The system is designed to process data from smartphones or wearable devices, making it flexible and suitable for various real-world applications such as fitness monitoring, elderly care, and smart home systems.
- **Comparative Evaluation**: The performance of the proposed model is compared against traditional machine learning models (SVM, KNN, Random Forest) to demonstrate the advantages of deep learning-based approaches in terms of accuracy and adaptability.
- **Real-Time Readiness**: The model is lightweight and optimized for edge deployment, allowing real-time recognition on mobile or embedded platforms like smartphones or Raspberry Pi.
- **Scalability and Usability**: The proposed framework can be extended to support new activity classes, user profiles, or additional sensor inputs with minimal reconfiguration, improving usability across domains.

Through this work, a robust, accurate, and scalable HAR system is developed that can operate in real-world environments while addressing key limitations of existing methods such as limited generalizability and reliance on static features.

III. SYSTEM ARCHITECTURE METHODOLOGY

The architecture of the Human Activity Recognition (HAR) system is structured as a modular and layered pipeline, enabling efficient real-time data handling—from collection to final activity classification. The system includes the following primary components:

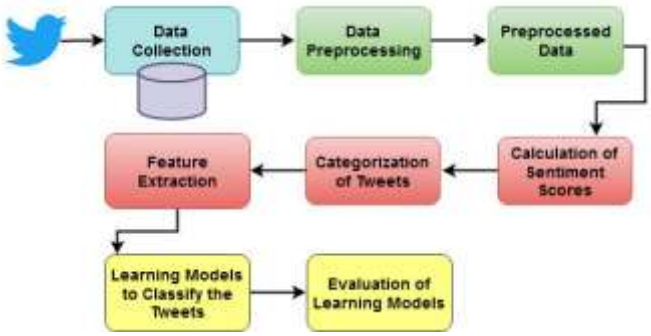
1. Data Collection Layer
   - Gathers raw sensor readings from devices such as accelerometers and gyroscopes embedded in wearables or smartphones.
   - Optionally integrates video input (e.g., camera frames) to support multimodal activity recognition using both sensor and visual data.
2. Preprocessing Module
   - Cleans and prepares the raw data by removing noise using filters such as Butterworth or median filters.
   - Performs signal normalization and segments the continuous stream into uniform time windows to ensure consistent input for the next stages.
3. Feature Extraction Layer
   - Employs Convolutional Neural Networks (CNNs) to identify spatial patterns in the sensor or image data.
   - Utilizes Long Short-Term Memory (LSTM) networks to capture and model the temporal dynamics and sequential behavior of activities over time.
4. Classification Module
   - Merges spatial features from CNNs with temporal patterns from LSTMs to accurately categorize activities (e.g., walking, standing, running).
   - A softmax layer is used at the final stage of the model to generate the probability distribution across predefined activity classes.
5. Output Interface
   - Displays the predicted activity labels in real time for user awareness or system logging.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

- Can be linked with broader applications such as healthcare systems, fitness monitoring apps, or personalized activity trackers.

IV. PROPOSED METHODOLOGY

The proposed methodology aims to develop an efficient and accurate Human Activity Recognition (HAR) system using a deep learning-based hybrid approach. The system is designed to process data from inertial sensors (accelerometer, gyroscope) collected through smartphones or wearable devices and classify human activities in real time.



4.1 User Upload Interface

The critical component that allows users to interact with the HAR system by uploading their sensor or video data for analysis. It is designed to be user-friendly, secure, and compatible with various data formats.

4.2 Video Output and Display

Real-Time Display

- **Input Source**: Live webcam feed or uploaded video.
- **Activity Overlay**: Detected activity (e.g., "Walking", "Sitting") is displayed as a text label on the video.
- **Confidence Score**: A confidence percentage is shown beside each predicted label.
- **Bounding Box (optional)**: If using pose detection or human detection, a box can highlight the person being analyzed.

Post-Processed Video Output

- **Video Playback**: The original or processed video is displayed with predicted activity labels synced with the timeline.
- **Color-coded Labels**: Different activities are represented with different colors for clarity.
- **Graph Overlay (optional)**: A real-time graph shows activity probabilities or sensor signals alongside the video.

V.DESIGN METHODOLOGY

5.1 Method Overview

The proposed Human Activity Recognition (HAR) approach is built around a deep learning framework that processes time-series sensor data—such as accelerometer and gyroscope readings—to classify different physical activities. The method integrates Convolutional Neural Networks (CNNs) to identify spatial characteristics in the data, and Long Short-Term Memory (LSTM) networks to model temporal patterns across time windows. By combining these two architectures, the model is capable of learning complex motion dynamics for accurate activity classification.

5.2 Pseudocode

BEGIN
1. Load Sensor Data
   - Import data from accelerometers and gyroscopes
   - Assign labels to corresponding activities (e.g., Sitting, Walking)
2. Data Preprocessing
   - Apply noise reduction techniques (e.g., low-pass filtering)
   - Normalize sensor readings

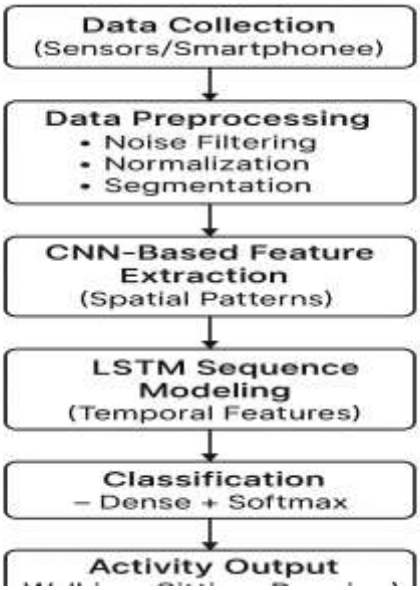
iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

- Segment the continuous data into equal-sized windows (e.g., 128 samples per segment)

3. Prepare Data for Training
   - Reshape into the format: (number of samples, time steps, features)
   - Convert labels to one-hot encoded format for classification
4. Design the CNN-LSTM Model
   - Input Layer: Accepts data of shape (time steps, features)
   - CNN Block:
     - Use 1D convolution layers to extract spatial features
     - Apply MaxPooling to reduce dimensionality
   - LSTM Block:
     - Feed CNN output into LSTM layers to learn sequential dependencies
   - Dense Layer: Fully connected for classification
   - Output Layer: Uses softmax activation to generate class probabilities
5. Compile the Model
   - Loss Function: Categorical Crossentropy
   - Optimizer: Adam
   - Evaluation Metric: Accuracy
6. Train the Model
   - Input training data into the model
   - Validate using a separate validation dataset
   - Train across multiple epochs with a defined batch size
7. Model Evaluation
   - Use the test dataset to make predictions
   - Calculate metrics such as Accuracy, Precision, Recall, and F1-score
8. Activity Prediction (Real-Time or Batch)
   - Feed new, preprocessed input window
   - Output the predicted activity label

END

5.3 FLOW DIAGRAM



iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

VI. IMPLEMENTATION

The Human Activity Recognition (HAR) system processes time-series data from sensors using a deep learning framework to classify various physical activities. The implementation is done using Python, leveraging powerful libraries like TensorFlow/Keras, NumPy, Pandas, and others. The system is capable of supporting both offline batch processing and real-time prediction modes.

6.1 SETUPS

1. Development Environment

- **Programming Language**: Python
- **Libraries & Tools**:
  - Deep Learning: TensorFlow, Keras
  - Data Handling: NumPy, Pandas
  - Visualization: Matplotlib
  - Machine Learning Utilities: Scikit-learn
- **Hardware Requirements**:
  - GPU-supported machine for training
  - Edge devices such as smartphones or Raspberry Pi for deployment

2. Data Collection

- The input data is sourced from publicly available datasets like UCI HAR or WISDM, or through custom collection using smartphone/wearable sensors.
- Each sample includes multi-axis readings from accelerometers and gyroscopes (X, Y, Z axes) and corresponding activity labels.

3. Data Preprocessing

- Apply filtering techniques (e.g., Butterworth filter) to reduce sensor noise.
- Normalize the readings to bring all values to a common scale.
- Segment the continuous data stream into fixed-size windows (e.g., 128 data points per segment) for uniform input.

4. Model Architecture

- The model follows a hybrid CNN-LSTM design:
  - CNN Layers: Extract spatial patterns from the sensor data.
  - LSTM Layer: Capture the temporal dependencies across time steps.
  - Dense Layer with Softmax: Classifies the activity into one of the predefined classes.

Python

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv1D, MaxPooling1D, LSTM, Dense
model = Sequential()
model.add(Conv1D(64, kernel_size=3, activation='relu', input_shape=(128, 6)))
model.add(MaxPooling1D(pool_size=2))
model.add(LSTM(100, return_sequences=False))
model.add(Dense(6, activation='softmax'))  # for 6 activity classes
```

5. Model Training and Validation

To build the Human Activity Recognition (HAR) model, the dataset is split into two parts: one for training and the other for validating the model's performance.

The model uses categorical crossentropy as the loss function, which is ideal for multi-class classification tasks involving various activity labels.

The Adam optimizer is selected for its adaptive learning capabilities, offering improved training speed and convergence stability.

Model performance is evaluated using a combination of metrics, including accuracy, precision, recall, and F1-score—ensuring a comprehensive assessment of how effectively the model distinguishes between activities.

6. Real-Time Activity Recognition (Optional)

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

For real-time use, the trained HAR model can be connected to continuous sensor input, such as data from a smartphone's accelerometer or a live video feed.
The system processes this data in real time and identifies the user's current activity on the fly.
Each prediction is shown immediately, often with an associated confidence score that reflects how sure the system is about the detected activity.

7. Output and Interpretation

The final output from the HAR system includes the predicted activity label—such as Walking, Sitting, or Running—along with a probability score that indicates the confidence level of the prediction.
The system may also log the recognized activities for future review or analytical purposes.
In applications such as fitness tracking or assistive technologies, results can be visualized dynamically, for example as on-screen overlays in a video stream or within a real-time dashboard or app interface.

6.2 Project Structure

Modular Design: Code is organized into independent modules for better maintainability and debugging.
Separation of Concerns: Different scripts handle model training, evaluation, and prediction tasks.
Scalable Architecture: Supports integration of additional datasets, models, and input types (such as video streams).
Portability: Compatible across platforms using virtual environments or Docker containers for easy deployment.

VII. CONCLUSION

In today's technology-driven world, rapid advancements in computer vision have made Human Activity Recognition (HAR) systems a critical and transformative innovation. These systems have demonstrated substantial effectiveness across a broad spectrum of real-world applications, ranging from security surveillance and behavioral monitoring to assisting the elderly and individuals with visual impairments.
The CNN-LSTM-based HAR system developed in this project reflects a significant advancement in the field. It provides a robust and accurate mechanism for identifying and classifying various human activities. One of its notable strengths lies in its flexibility to process both video streams and image-based input, making it well-suited for integration into diverse industrial and consumer applications.
Looking ahead, there is enormous potential to further improve and expand HAR systems. Emerging research areas include:

- Enhancing model accuracy through advanced architectures (e.g., transformers or graph neural networks),
- Integrating HAR into IoT and edge devices for real-time, low-latency applications,
- Increasing robustness across varied environments and diverse populations,
- Incorporating multimodal inputs (e.g., audio, depth, and environmental data) for better contextual understanding.

VIII. REFERENCES

1) J. Zhang, A. Yang, C. Miao, X. Li, R. Zhang and D. N. H. Thanh, "3D Graph Convolutional Feature Selection and Dense Pre-Estimation for Skeleton Action Recognition," in IEEE Access, vol. 12, pp. 11733-11742, 2024, doi: 10.1109/ACCESS.2024.3353622.
2) [M. Atikuzzaman, T. R. Rahman, E. Wazed, M. P. Hossain and M. Z. Islam, "Human Activity Recognition System from Different Poses with CNN," 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2020, pp. 1-5, doi: 10.1109/STI50764.2020.9350508.
3) M. Ramzan, A. Abid, M. Bilal, K. M. Aamir, S. A. Memon and T. -S. Chung, "Effectiveness of PreTrained CNN Networks for Detecting Abnormal Activities in Online Exams," in IEEE Access, vol. 12, pp. 21503-21519, 2024, doi: 10.1109/ACCESS.2024.3359689.
4) S. Win and T. L. L. Thein, "Real-Time Human Motion Detection, Tracking and Activity Recognition with Skeletal Model," 2020 IEEE Conference on Computer Applications(ICCA), Yangon, Myanmar, 2020, pp. 1-5, doi: 10.1109/ICCA49400.2020.9022822.
5) A. Gupta, K. Gupta, K. Gupta and K. Gupta, "A Survey on Human Activity Recognition and Classification," 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2020, pp. 0915-0919, doi: 10.1109/ICCSP48568.2020.9182416.