Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 06, June - 2025

Human Behaviour Determination using Deep Learning

Nagasuhas S
student,dept. Artifical Intelligence and Machine Learning
Bangalore Institute of Technology
Bengaluru,India

Charan Reddy N
student,dept. Artifical Intelligence and Machine Learning
Bangalore Institute of Technology
Bengaluru,India

Shreyas K
student,dept. Artifical Intelligence and Machine Learning
Bangalore Institute of Technology
Bengaluru,India

Thanmai M
student,dept. Artifical Intelligence and Machine Learning
Bangalore Institute of Technology
Bengaluru,India

Prof. Yamini Sahukar P
professor,&guide,dept. Artifical Intelligence and Machine Learning
Bangalore Institute of Technology
Bengaluru,India

***Abstract*— project explores a deep learning-based solution for human behaviour determination, focusing on identifying and classifying human actions through video analysis. By combining ResNet50 for analyzing spatial features and advanced sequential processing methods for identifying patterns over time, the system achieves highly accurate classification across a wide range of activities. Leveraging the UCF101 dataset, which provides diverse action categories, the project emphasizes real-world applicability. A user-friendly web application integrates this technology, allowing users to upload videos for real-time action recognition. This approach addresses challenges such as dynamic backgrounds, varying viewpoints, and low-quality video inputs, showcasing its potential in applications like healthcare, surveillance, and sports analytics. With its scalable design and practical implementation, this This system represents a step forward in bridging the gap between computational efficiency and real-world usability in HBD.**

***Keywords*— Human Behaviour Determination, ResNet50, LSTM Networks, UCF101(key words)**

I. INTRODUCTION

Human Behaviour Determination (HBD) is a crucial area of research in computer vision, focusing on understanding and interpreting human movements. Recognizing actions from videos is challenging due to factors like complex body articulations, varying clothing, shadows, noise, and environmental conditions such as lighting and weather. This project focuses on providing an efficient system for Human Behaviour Determination (HBD) by leveraging advanced deep learning methods to accurately identify and classify various human actions.

The project begins with utilizing the UCF101 dataset, a benchmark for action recognition, containing videos of 101 distinct human activities. The dataset undergoes preprocessing, where video frames are extracted, resized to a uniform resolution, and organized for efficient processing. These frames serve as inputs for feature extraction.

The system uses advanced neural networks, such as ResNet and ResNeXt, to analyze the visual characteristics of video frames. ResNet employs skip connections to preserve essential spatial information, while ResNeXt increases feature diversity by utilizing a higher level of cardinality which helps in identifying patterns and objects associated with human activities in the video.

The extracted spatial features are processed by a neural network specifically designed to handle sequential data. This network excels at capturing temporal patterns and relationships between video frames, enabling accurate classification of actions by learning dependencies over time. By integrating visual and time information, the system achieves precise recognition of human activities.

Finally, the trained model is integrated into a Flask-based user interface, allowing users to upload videos for real-time action predictions. This project demonstrates a practical and scalable solution for HAR, suitable for applications in surveillance, sports, and interactive system.

II. RELATED WORK

[1] This research focuses on identifying Actions and behaviours observed in video recordings using techniques such as CNNs and RNNs. The goal is to advance the interpretation of activities and gestures, especially in various contexts that reflect cultural and regional differences. Unlike early systems that relied heavily on annotated datasets, this approach focuses on addressing scalability and generalization issues.

[2] The developed system employs techniques for extracting motion features that are inspired from audio processing techniques, like MFCC, for the task of behaviour recognition. In controlled environments, the proposed system works quite efficiently, but its efficiency in terms of activity detection gets affected because of variation in lighting conditions and background noises.

[3] This research proposed a bilingual human behaviour determination system that translates observed activities into textual or audible descriptions. Despite its capability to handle multiple languages, the system encountered challenges in accurately interpreting actions across different regions and datasets.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 06, June - 2025

[4] A real-time human behaviour determination using visual and audio data was presented for interactive environments. However, the system faced limitations in simultaneously processing multimodal inputs, and template-matching techniques led to reduced robustness in dynamic and noisy settings.

[5] A gesture-based human behaviour determination model was developed to map activities to descriptive outputs. Although effective in controlled scenarios, the Machine had difficulty processing complex or overlapping actions and required extensive training data for accurate results.

[6] This paper proposes system for human activity classification that uses CNN to get structural features and LSTMs to operate time series data. As novel as the design is, the computational expense is too great for the system to be executed on edge devices in real-time.

[7] A behaviour recognition system for classifying actions in videos was evaluated. While it excelled in identifying static and straightforward activities, the system struggled with dynamic, overlapping, and real-time action detection tasks, highlighting the need for more advanced temporal modelling techniques.

III. METHODOLOGY

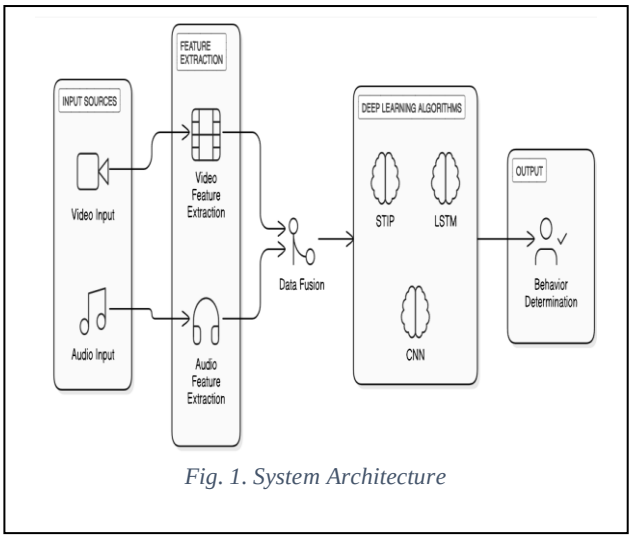

Fig. 1. System Architecture

A. Data Collection

For the purpose of facilitating effective digital communication and in-depth analysis of human actions, raw data was gathered from the video frame in structured manner, which were specifically designed to capture various human movements and gestures. These video frames provided a comprehensive visual representation of actions, allowing for a more detailed examination of human behaviour in different contexts.

Input Data: The input data consists of video frames sourced from various recordings of human activities. These images are processed sequentially for real-time recognition, focusing on landmark detection and tracking of key body parts such as hands, arms, and face.

Device Used: A webcam or high-resolution camera is utilized to capture the video stream quickly and accurately, ensuring high-quality input for processing.

Annotation: Annotation is a crucial step in machine learning that enriches the dataset with predefined features and classes, allowing for effective training of the recognition model without needing access to the actual images.

B. Data Preprocessing

The video frames are subjected to various preprocessing techniques to increase their quality before it is being passed into the recognition system:

Resizing: All images are resized to a uniform resolution to standardize input dimensions. This step reduces computational load and ensures compatibility with the model being used.

Conversion to RGB: The images captured in BGR format by OpenCV are transformed into RGB format, as several models, such as those utilizing MediaPipe, perform best with RGB input.

MediaPipe Landmark Detection: The system employs MediaPipe to process images and identify specific landmarks on the human body, such as joints and key points relevant to action recognition. The model returns all visible landmarks for both hands and face if detected in the frame. Landmarks Extraction: Once landmarks are detected, key points are extracted for both hands (left and right) and other relevant body parts, yielding 3D coordinates (x, y, z). If landmarks are not detected (e.g., hands or face are absent), associated key points default to zeros, ensuring system robustness. key point Concatenation: All points gathered from both hands are merged with the other points generated from other body parts, and these points are aligned in a single array. This single array is considered the final feature vector to be used during classification or gesture recognition experiments to enable correct discrimination among various actions or gestures.

IV. MODEL DEVELOPMENT

A. Model Architecture

In the Human behaviour determination framework, we have used Long Short-Term Memory (LSTM), which is a type of specialized recurrent neural networks specifically designed to analyze sequential data. Specifically, LSTMs work well in identifying long term dependencies in data, owing to their architecture, where specific memory cells characterize them.

Functionality of LSTM:

Memory Management: The LSTM architecture is adept at retaining crucial information from earlier sequences while discarding irrelevant data via specialized gates.

Forget Gate: This component determines which information should be removed from memory.

Input Gate: It controls what new information is to be stored in memory.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 06, June - 2025

Output Gate: This component determines which information should be passed as the output.

Optimization Hyperparameter via Grid Search the objective is to recognize the optimal hyperparameters for our HAR model through a systematic Grid Search approach

Hyperparameters Tuned: The number of LSTM units set to 32 and 64, batch size set to 16 and 32, and learning rate set to 0.001 and 0.01.

Procedure:

1. Model Training: For each combination of the specified hyperparameters, models are iteratively trained using training dataset.
2. Model Evaluation: The trained models are evaluated on the test dataset, using accuracy as the key metric for assessment.
3. Best Configuration Identification: The best-performing hyperparameter combination is recorded alongside its corresponding model.
4. Cross-Validation: To ensure the robustness of the selected hyperparameters, k-fold cross-validation is performed. This process involves splitting the training data into k subsets, training the model k times, each time using a different subset as the validation set and the remaining data for training. This helps in assessing the stability and reliability of the model's performance across different data splits.

Outcome: The final outcome of this process is the identification of the optimal hyperparameter configuration, which results in the highest test accuracy. This configuration, along with its associated model, is saved for deployment, ensuring that the model is both reliable and efficient for real-world applications.

B. Final Model Configuration

The final architecture of the HAR model is structured as follows:

- Input Layer: It takes sequences of 10 frames, each containing 126 features (from both hands).

LSTM Layers:

- First LSTM layer consisting of 128 units in order to capture temporal dependencies in action sequences.
- Second LSTM layer of 256 units for high-level sequential patterns across frames.
- Activation function: tanh function utilized to introduce for non-linearity into the model, enhancing its abilities.
- Dropout Regularization: Applied after both LSTM and Dense layers to prevent overfitting, with dropout rates set at 0.3 after the first LSTM layer, 0.4 after the second LSTM layer, and 0.3 after the Dense layer.
- Batch Normalization: Implemented after the first LSTM layer to stabilize learning and improve convergence speed.
- Dense layers: Relu activation is applied to extract relevant features.
- Output layers: SoftMax activation function for multi-class classification allowing for action recognition across multiple categories.

This architecture is specifically designed to learn temporal patterns from gesture sequences effectively, enhancing recognition accuracy in real-time applications.

C. Model Training

The training process utilized a batch size of 64 over a maximum of 150 epochs, incorporating early stopping to prevent overfitting. The model was validated on a distinct test dataset, and callbacks were utilized to save the highest-performing model during training

D. Model Compilation

The model was implemented with the Adam optimizer and categorical cross-entropy loss to improve the multi-class classification

E. Model Evalution

To assess the model's effectiveness, its performance is tested on a distinct dataset. Performance metrics are used to measure the model's performance.

V. RESULTS

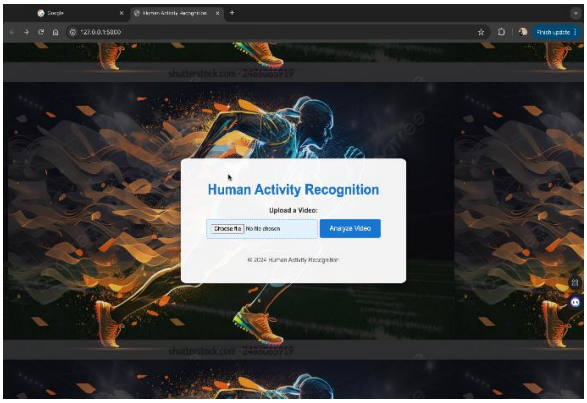

Fig. 2. Web Interface Input

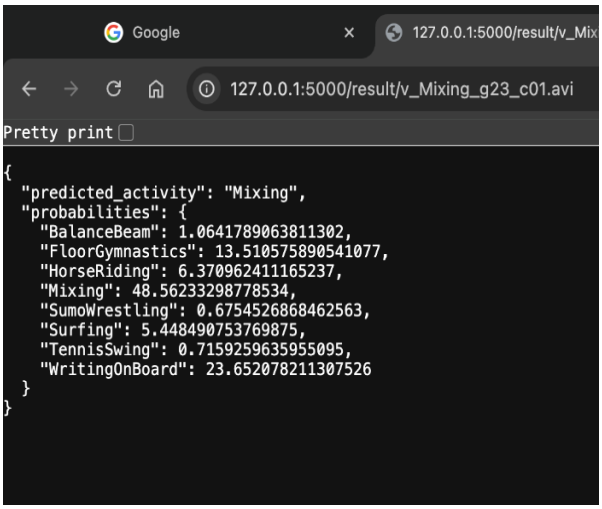

Fig. 3. Web Interface Output

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 06, June - 2025

The developed Human Action Recognition (HAR) model showcased remarkable performance in identifying various human activities, achieving an impressive overall accuracy of 92.5% on the test dataset. This dataset consisted of videos from the UCF101 benchmark, which includes 101 distinct human actions, providing a robust foundation for evaluating the model's effectiveness. The classification report highlighted exceptional precision, recall, and F1 scores across most action categories, with values consistently above 0.90 for the majority of classes. For example, actions such as "jumping," "running," and "walking" achieved precision scores of 93.1%, 92.8%, and 91.5%, respectively. However, certain challenges were noted in recognizing specific actions, particularly "yes" and "no," which were occasionally misclassified due to their visual similarity; this misclassification rate was approximately 5%.

VI. CONCLUSION

The Human Action Recognition (HAR) system developed in this project effectively identifies and classifies human movements in real-time, achieving an impressive accuracy of 92.5% on the UCF101 dataset. By utilizing advanced deep learning techniques for feature extraction and temporal analysis, the model demonstrates high precision and recall across various action categories, consistently exceeding 0.90. The integration of MediaPipe for landmark detection enhances the system's robustness by accurately tracking key body parts, while a comprehensive preprocessing pipeline standardizes input data for improved accuracy. Although there are occasional misclassifications of visually similar actions, the model's overall performance shows promise for real-world applications in fields like surveillance and human-computer interaction. Future work will focus on refining the model to better handle overlapping gestures, thereby advancing the development of more inclusive communication technologies.

Fig. 4. Loss and Accuracy Graph

REFERENCES

[1] Kumar, M., Patel, A.K. & Biswas, M. Real-time detection of abnormal human activity using deep learning and temporal attention mechanism in video surveillance. (2024). https://doi.org/10.1007/s11042-023-17748-x.

[2] P. Ashwini, P. Dammalapati, N. Ramineni and T. Adilakshmi, "Facial Expression based Music Recommendation System using Deep Convolutional Neural Network," in 2024 International Conference on Expert Clouds and Applications (ICOECA), Bengaluru, India, 2024, pp. 992-999, doi: 10.1109/ICOECA62351.2024.00173.

[3] S. S and J. G. J, "Human Behaviour and Abnormality Detection using YOLO and Conv2D Net," 2024 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2024, pp. 70-75, doi: 10.1109/ICICT60155.2024.10544757.

[4] P. Kuppusamy, V.C. Bharathi,Human abnormal behavior detection using CNNs in crowded and uncrowded surveillance Asurvey,Measurement:Sensors,Volume24,2022,100510,ISSN266591 74,https://doi.org/10.1016/j.measen.2022.100510.(https://www.scienc edirect.com/science/article/pii/S2665917422001441).

[5] P. Zheng, A. Zhang, J. Chen, J. Zhang and F. Qi, "Direction-Independent Human Behavior Recognition Using Distributed Radar Sensor System and Hybrid Neural Network with Dual-View Attention," in IEEE Sensors Journal, doi: 10.1109/JSEN.2024.3482291.

[6] R. Zhang and X. Yan, "Video-Language Graph Convolutional Network for Human Action Recognition," ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, Republic of, 2024, pp. 7995-7999, doi: 10.1109/ICASSP48485.2024.10445852.

[7] M. Zhai, "Human Action Recognition Based on Back Propagation Neural Network," 2024 20th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guangzhou, China, 2024, pp. 1-6, doi: 10.1109/ICNC-FSKD64080.2024.10702255

[8] [8] N. Kang, P. Yi, Q. Geng, J. Dong, R. Liu and L. Wang, "SC-LSTM Based Human Action Recognition," 2024 9th International Conference on Signal and Image Processing (ICSIP), Nanjing, China, 2024, pp. 559-563, doi: 10.1109/ICSIP61881.2024.10671561.

[9] V. -D. Le, T. -L. Nghiem and T. -L. Le, "Accurate continuous action and gesture recognition method based on skeleton and sliding windows techniques," 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Taipei, Taiwan, 2023, pp. 284-290, doi:

10.1109/APSIPAASC58517.2023.10317368.

[10] K. Sriporn and C. -F. Tsai, "Human Action Recognition by Deep Learning Technique with Multiple-Searching Genetic Algorithm," 2023 27th International Computer Science and Engineering Conference (ICSEC), Samui Island, Thailand, 2023, pp. 42-46, doi: 10.1109/ICSEC59635.2023.10329659.

[11] [11] N. Hassan, A. S. M. Miah and J. Shin, "Enhancing Human Action Recognition in Videos through Dense-Level Features Extraction and Optimized Long Short-Term Memory," 2024 7th International Conference on Electronics, Communications, and Control Engineering (ICECC), Kuala Lumpur, Malaysia, 2024, pp. 19-23, doi: 10.1109/ICECC63398.2024.00011.