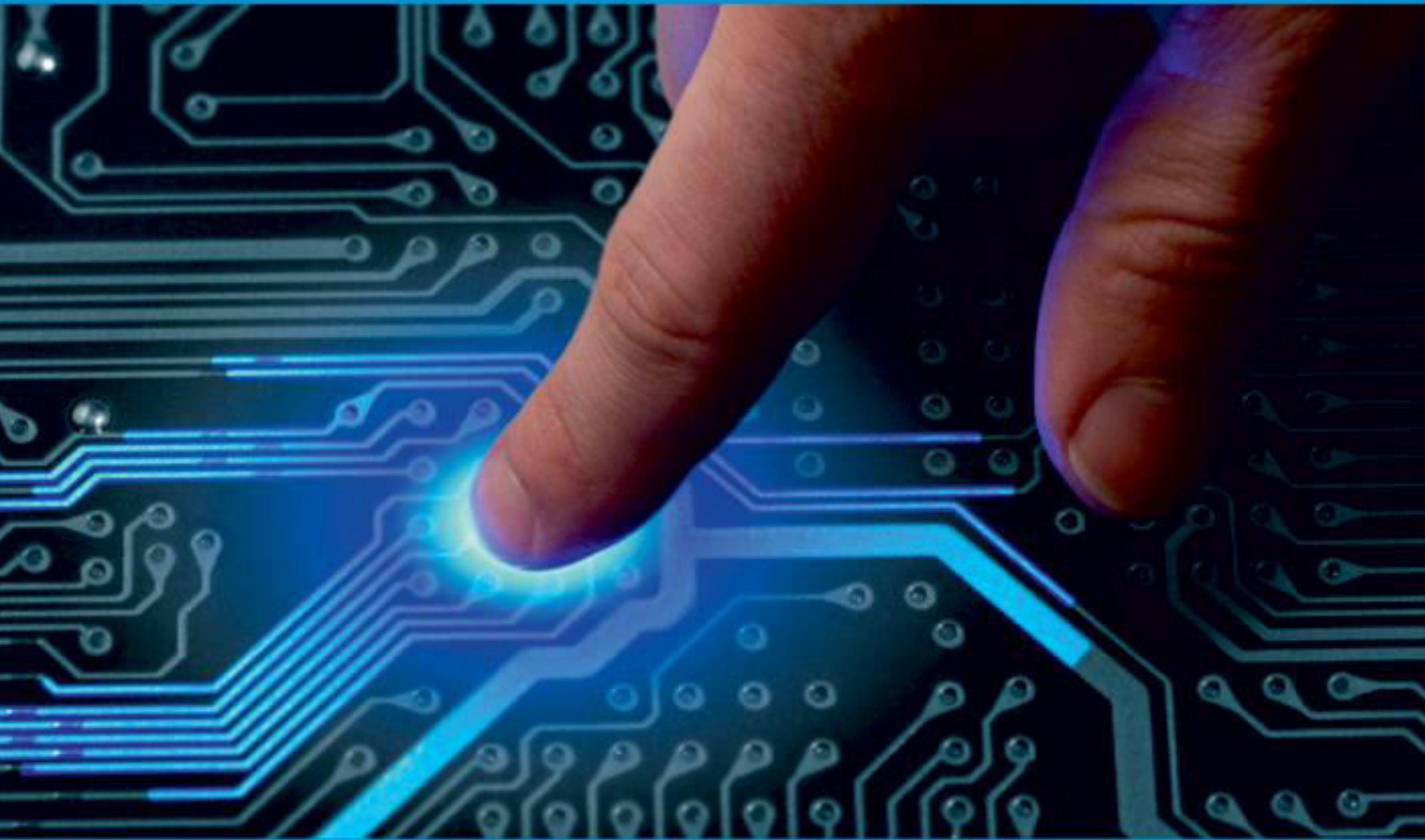
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 12, December 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

© 2024 IJIRCCE | Volume 12, Issue 12, December 2024| DOI: 10.15680/IJIRCCE.2024.1212059

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Deep Learning for Human Activity Recognition: A Focus on Walking and Running

M Surekha, G.Jyothirmai

M. Tech, Department of CSE, St. Mary's Women's Engineering College, Guntur, A.P., India

Assistant Professor, St. Mary's Women's Engineering College, Guntur, A.P., India

ABSTRACT: This project addresses human action detection in dynamic environments by proposing a novel approach that integrates traditional temporal modeling with modern deep learning techniques. Our method utilizes a multi-task framework, combining classification and regression objectives to identify action types and precisely localize their temporal boundaries. Key components include:

Recurrent Neural Networks (RNNs): To capture long-range dependencies and dynamic motion patterns.
Temporal Constraint Representation: Inspired by interval algebra, to handle noisy or imprecise annotations robustly.
Graph Convolutional Networks (GCNs): To reason about interactions between actors and their environments, enhancing context-aware detection.

Evaluation on public datasets demonstrates superior performance in both classification accuracy and temporal localization compared to baseline methods. This work bridges the gap between traditional temporal constraint networks and state-of-the-art deep learning paradigms.

I. INTRODUCTION

Human action detection in videos is crucial for applications like surveillance and human-computer interaction. Existing methods face challenges such as:

Complex temporal structures
Cluttered backgrounds
Ambiguous spatial-temporal boundaries
Computational overhead
Limited generalization
This research aims to address these challenges by developing a robust and efficient action detection system that:

Accurately identifies action types and localizes their temporal boundaries.
Handles complex scenarios with cluttered environments and diverse viewpoints.
Enables real-time processing for practical applications.

Input file data	Name of dataset	File format	File size	No.sample
Videos performing different actions	HMDB datase UCF101data set	.avi .avi	528 **kb** 256 **kb**	50405 13456

This condensed version maintains the core message while significantly reducing the text length.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. KEY CHANGES

Removed redundant phrases and sentences.
Combined similar ideas for better conciseness.
Used bullet points to enhance readability.
Focused on the key contributions and novel aspects of the proposed method.

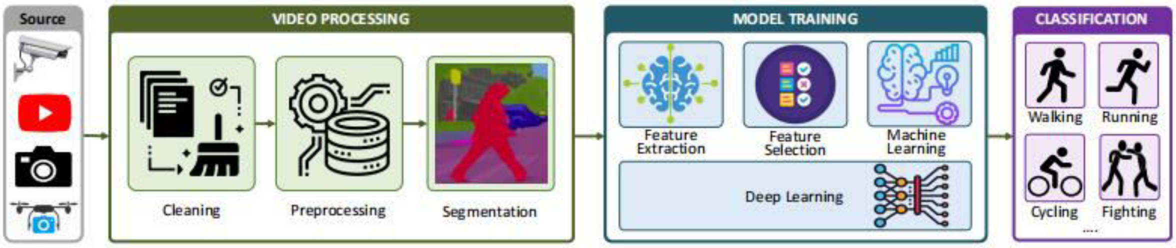


Figure 1. Human Action Recognition Framework.

III. RESEARCH METHOD AND TAXONOMY

To provide a thorough and up-to-date review of human action recognition (HAR), we employed a structured and detailed research methodology. The approach involved the following steps:

Defining the Scope and Objectives: We began by clearly defining the goals and scope of the study. This involved outlining the key aspects of HAR that would be examined, including its historical development, recent advancements, and current state.

Conducting a Comprehensive Literature Search: A thorough search of the academic literature was conducted using various sources such as Google Scholar, MDPI, PubMed, and IEEE Xplore. The aim was to identify relevant studies, articles, and publications that contribute to the field of human action recognition.

Evaluating the Quality of the Literature: Each piece of literature was critically assessed based on several factors, including the validity and reliability of the research methods, the alignment of the findings with the study's objectives, and the quality of data analysis.

Classifying the Literature: The collected literature was organized into categories based on the components of HAR being examined. This included methods focused on feature extraction, activity types, and other relevant classifications.

Synthesizing the Literature: We synthesized the findings by summarizing the key points of each study, comparing different methodologies and results, and providing original insights based on the analysis.

Analyzing and Interpreting the Data: The data collected from the literature was analyzed to address specific issues within the field of HAR. The analysis aimed to draw conclusions, identify gaps in existing research, and suggest directions for future studies.
This methodical approach enabled the development of a comprehensive and up-to-date review of human action recognition. The aim was to provide meaningful insights into this rapidly evolving area of research.

IV. ACTION CLASSIFICATION LEVELS

In the current study, action classification is explored at four semantic levels: atomic, behavior, interaction, and group. The focus has primarily been on the first two levels—atomic actions and behavior—while research on recognizing group activities remains limited. Despite growing interest in interaction recognition, group-level activity recognition has not been widely explored within the research community. Moreover, the representation of action features remains a

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

core issue, particularly for basic categorization and individual actions. Much of the current research on HAR concentrates on recognizing fundamental actions performed by a single person.

This survey provides an in-depth review of these areas, shedding light on the ongoing research efforts and identifying potential areas for further exploration in the field of human action recognition.

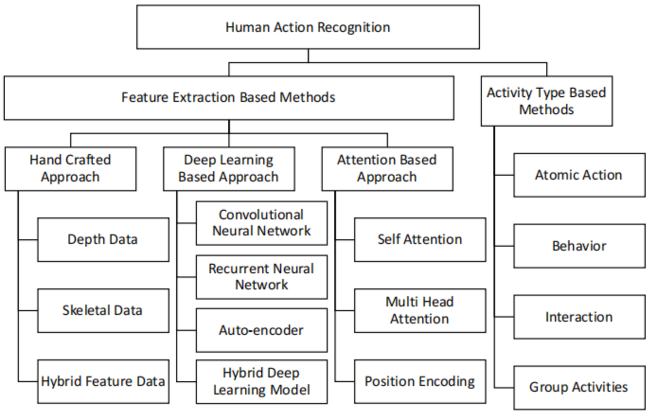


Figure 2. Human Action Recognition Taxonomy

V. FEATURE EXTRACTION-BASED ACTION RECOGNITION

In this project, the approach to feature extraction for human action recognition is based on processing video data from the HMDB dataset. The process involves extracting individual frames from the videos, resizing, and normalizing them to prepare them for input into a model. This ensures that the model receives uniform data for training, enabling it to learn to classify human actions accurately.

5.1 Data Loading and Preprocessing

The load_hmdb_data_in_batches function is used to load the videos from the HMDB dataset in batches. Each video is read frame by frame, and the frames are resized to a target image size (e.g., 64x64). These frames are normalized by dividing by 255 to ensure that pixel values are in the range [0, 1]. A sequence of frames (defined by sequence_length) is then created from each video. This sequence is important for recognizing dynamic actions over time, as human activities are typically temporally dependent.

This function is capable of processing the videos in batches, which helps manage memory usage and speeds up the data preparation process. The sequences are stored as NumPy arrays and returned in batches for further processing or model training.

5.2 Frame Extraction

To extract frames from a video, the frames_extraction function is used. It reads a video file, extracts each frame, resizes it to the desired dimensions (128x128 in this case), and normalizes the pixel values. These frames are the primary features used in training the action recognition model.

5.3 Dataset Creation

After extracting frames, the create_dataset function assembles the training and testing datasets. It takes in the list of class names and their corresponding video files, extracts frames, and assigns labels based on the class. The dataset is divided into training and testing sets, and a subset of frames (randomly selected) is used to ensure diversity in the dataset.

5.4 Data Augmentation

To enhance the model's ability to generalize, data augmentation techniques (like randomly sampling frames from videos) are applied. This technique helps prevent overfitting by providing the model with varied training examples.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

To prevent overfitting, we apply data augmentation to create more varied training examples. This can be done by randomly sampling frames from the videos. Let n be the number of frames in a video and m the number of frames randomly selected for each sequence:

$$\text{Augmented Sequences} = \binom{n}{m}$$

Where (M/n) represents the number of ways to randomly select m frames from n. This technique ensures that the model learns generalizable patterns and not just memorizing sequences.

5.5. Loss Function for Action Recognition

For action recognition, a common loss function used in deep learning models is **categorical cross-entropy** for multi-class classification. If represents the true label and $\hat{y}$ represents the predicted output, the categorical cross-entropy loss L is given by:

Where:

C is the number of classes (e.g., walk, run, etc.).

y_i is the true label for class iii, which is 1 for the correct class and 0 for others.

$\hat{y}_i$ is the predicted probability for class iii from the model's output.

5.6 Final Dataset

After data extraction and augmentation, the final dataset consists of image sequences, each labeled according to the action being performed. These sequences are used to train machine learning or deep learning models for action recognition. In this case, the action categories include "walk" and "run."

Train Batch={X_{train}, Y_{train}}for i=1,2,...,num_batches

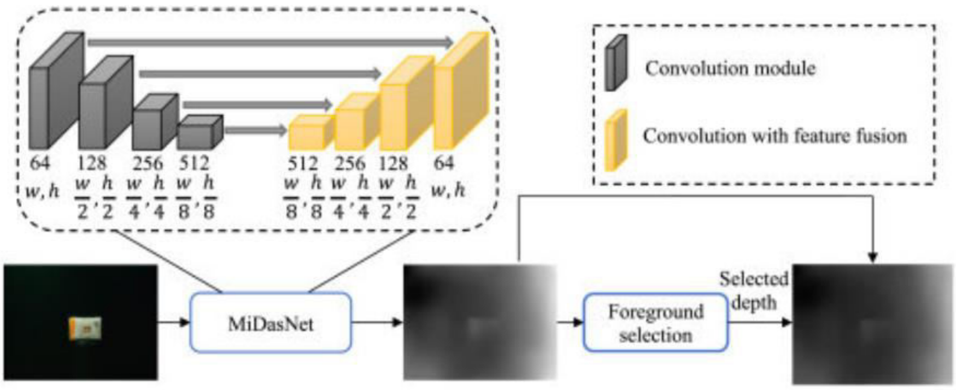


Figure 3:Depth-Based Approaches

The 3D points obtained from surface image frames can be utilized to calculate normal vectors, which help in extracting motion and shape features for action recognition models [32, 33, 34]. In [19], a new descriptor for action recognition using depth sequences was introduced. This descriptor effectively captures both shape and motion information from a

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4D space normal orientation histogram, incorporating depth images, time, and coordinates. In [33], a super normal vector was proposed, utilizing polynormal encoding to represent local shape and motion data. Slama et al. [34] introduced a framework that modeled local displacement features as subspaces on a Grassmann manifold and used them to create a probability density function for action categorization.

VI. SKELETON-BASED METHODS

Skeleton-based methods offer an alternative way to extract information about the human body using depth measurements, as shown in Figure 6. The low-dimensional nature of skeleton data [21] enables faster execution of Human Activity Recognition (HAR) models. Using depth cameras to generate a 3D representation of human joints is a promising research direction with numerous potential applications. Representing human actions through the body skeleton remains an open area of exploration among researchers. The human body's joints can be represented by 2D or 3D coordinates extracted from depth images, which can then be tracked to capture motion features.

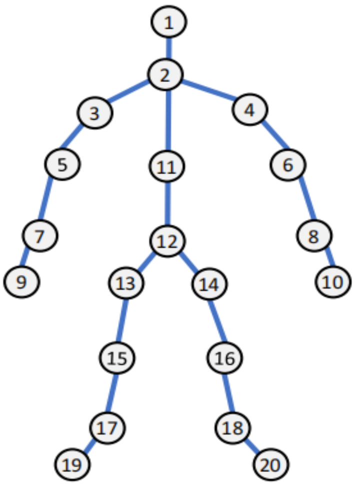


Figure 4:Skeleton model created with Kinect sensor in 3D. The dots indicate 20 joints, while the lines represent 19 limbs

VII. DEEP LEARNING REPRESENTATION METHOD

PoseCOV3D is a technique designed to model human poses in 3D space over time, capturing both spatial and temporal relationships effectively. It leverages covariance descriptors to represent pose trajectories in 3D space, making it highly effective for action recognition tasks.

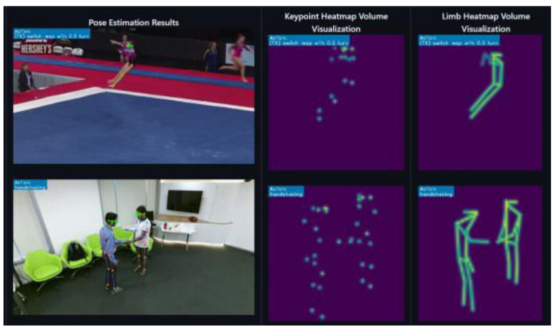


Figure 5:Skeleton model created with Kinect sensor in 3D. The dots indicate 20 joints, while the lines represent 19 limbs

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

PoseCOV3D is a technique designed to model human poses in 3D space over time, capturing both spatial and temporal relationships effectively. It leverages covariance descriptors to represent pose trajectories in 3D space, making it highly effective for action recognition tasks.

7.1 Steps to Implement PoseCOV3D:

The PoseCOV3D approach begins with extracting 3D poses from video data using pre-trained pose estimation libraries such as OpenPose or MediaPipe, which provide 3D joint coordinates for each frame. These extracted poses are then processed to compute covariance descriptors, representing spatial relationships between joint positions over time through covariance matrices. These descriptors serve as the primary feature representation for each action sequence, effectively encapsulating motion and spatial information. Finally, these covariance-based features are classified into action categories using machine learning models such as SVMs, LSTMs, or fully connected networks, enabling precise and efficient action recognition.

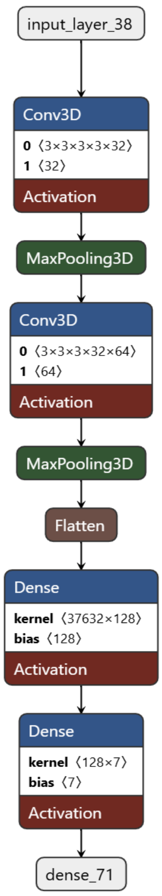


Figure 6: model digram of 3d cnn model

VIII. EVALUATION METRICS AND PERFORMANCE IN HUMAN ACTIVITY RECOGNITION

In the realm of human activity recognition, accurately assessing the performance of classification models is paramount for ensuring their reliability and effectiveness in real-world applications. A suite of metrics, adapted and refined from general classification tasks, provides valuable insights into a model's predictive capabilities. These metrics go beyond simple accuracy to offer a nuanced understanding of strengths and weaknesses, guiding researchers towards continuous improvement.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

8.1Core Metrics:

Precision: This metric focuses on the accuracy of positive predictions. It measures the proportion of correctly identified instances within the set of instances predicted to belong to a specific class. High precision indicates fewer false positives, minimizing the risk of incorrect alarms or misinterpretations.

$$\text{Precision} = Tp/(Tp + Fp)$$

8.2 Recall (Sensitivity): Recall measures the model's ability to identify all relevant instances within a given class. It quantifies how many actual positives the model successfully detected. High recall is crucial in scenarios where missing instances have significant consequences, such as in medical diagnosis or security applications.

$$\text{Recall} = Tp/(Tp + Fn)$$

8.3 F1-Score: The F1-score provides a balanced measure that considers both precision and recall. It's particularly valuable when dealing with imbalanced datasets, where a single metric might not accurately reflect overall performance. A high F1-score signifies a good balance between precision and recall, indicating a model that effectively identifies relevant instances while minimizing false positives

$$\mathbf{x} = \frac{\mathbf{2} * \mathbf{Recall} * \mathbf{Precision}}{\mathbf{Recall} + \mathbf{Precision}}$$

8.4 Accuracy: This is the most straightforward metric, representing the overall proportion of correct predictions across all classes. While easy to interpret, accuracy can be misleading in imbalanced datasets where the majority class dominates.

$$\mathbf{x} = \frac{\mathbf{Number\ o\ f\ Correct\ Predictions}}{\mathbf{Total\ Number\ o\ f\ Predictions}}$$

8.5 Confusion Matrix: The confusion matrix offers a comprehensive visualization of classification outcomes for each class. It presents a tabular representation of true positives, true negatives, false positives, and false negatives, providing insights into the types of errors the model makes. This granular information is invaluable for identifying misclassification patterns and fine-tuning the model.

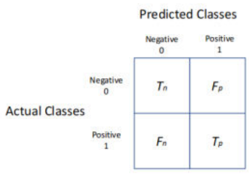


Figure 7:Confusion Matrix

Advanced Metrics:

Matthews Correlation Coefficient (MCC): MCC provides a balanced measure that accounts for true positives, true negatives, false positives, and false negatives. It ranges from -1 to +1, with +1 indicating perfect prediction and -1 representing complete disagreement. MCC is particularly useful in imbalanced datasets as it considers the balance between classes.

Beyond Basic Metrics:

Class-Specific Metrics: In multi-class classification, it's crucial to evaluate performance on each class individually. This can reveal class-specific biases or challenges, guiding targeted improvements.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Time-Series Considerations: For time-series data, metrics that account for temporal dependencies, such as dynamic time warping distances or segment-level evaluation, may be more appropriate.
Interpretability Metrics: In addition to predictive accuracy, understanding the rationale behind a model's predictions is crucial. Metrics related to model interpretability, such as feature importance or rule extraction, can provide valuable insights.

ACTION TYPE		ACCURACIES	METHOD	YEAR
Atomic Action	KTH	99.86%	PredRNN-v2[51]	2021
	NTURGB+D	97.1%	PoseC3D [52]	2022
	MSR Action 3D	98.02%	Temporal Subspace Clustering [53]	2021
Behavior	MCAD	86.9%	ConfluxL STMs network [54]	2021
Interaction	MSR Daily Activity	97.5%	DSSCA-SSLM [55]	2017
	3D	100%	ST-LSTM (Tree) + Trust	2016
	MuHAVI	94.4%	Gate [56]	2015
	UCF50		MIFS [57]	

IX. RESEARCH ISSUES

Data Acquisition Challenges:
Lighting Variations: Changes in lighting conditions can drastically affect image quality, making it difficult for visual-based systems to accurately recognize actions.
Viewpoint Limitations: Systems relying on single-viewpoint cameras are restricted in their ability to observe actions from different perspectives, leading to potential misinterpretations.
Occlusion: Self-occlusion (body parts blocking each other), occlusion by objects in the environment, and partial occlusions of body parts pose significant challenges, especially for systems relying on visual data.
Sensor Noise and Drift: Sensor-based systems can be affected by noise and drift in sensor readings, leading to inaccurate data and unreliable activity recognition.

Data Representation Challenges:
High Dimensionality: Raw sensor data and video streams often have high dimensionality, making it computationally expensive to process and analyze.
Complex Action Representations: Representing complex human actions, such as those involving multiple people or interactions with objects, can be challenging.
Intra-class Variability: Significant variations in how individuals perform the same action (e.g., walking styles) can make it difficult to accurately classify activities.
Inter-class Similarity: Some activities may exhibit subtle differences, making it challenging to distinguish between them.

Computational Challenges:
Real-time Processing: Many HAR applications require real-time processing, which can be computationally demanding, especially for complex models and high-resolution data.
Resource Constraints: Deploying HAR systems on resource-constrained devices, such as wearable sensors and mobile devices, requires efficient algorithms and optimized implementations.

Data Limitations:
Limited Datasets: Many publicly available datasets are relatively small and may not adequately represent the diversity of human activities and environments.
Data Bias: Datasets may exhibit biases in terms of the demographics of the subjects, the types of activities represented, and the environmental conditions.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Labeling Challenges: Accurately labeling large datasets of human activities can be time-consuming and expensive, requiring expert human annotators.

Opportunities
Real-World Applications:
Healthcare: Monitoring patient activity levels, detecting falls, and providing personalized healthcare interventions.
Smart Homes: Enabling intelligent home environments that can adapt to the needs and preferences of residents.
Surveillance and Security: Enhancing security systems by detecting suspicious or abnormal behavior.
Human-Computer Interaction: Improving the naturalness and intuitiveness of human-computer interfaces.
Robotics: Enabling robots to better understand and interact with humans in various social and work environments.

Technological Advancements:
Deep Learning: Leveraging deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to extract high-level features from raw data and improve recognition accuracy.
Multi-modal Approaches: Combining data from multiple sensors (e.g., cameras, microphones, wearable sensors) to improve robustness and accuracy.
Edge Computing: Processing data locally on edge devices to reduce latency, improve privacy, and conserve bandwidth.

X. FUTURE DIRECTIONS

Developing Robust and Generalizable Models:
Domain Adaptation: Developing models that can adapt to new environments and unseen activities.
Unsupervised and Semi-supervised Learning: Exploring techniques that can learn from unlabeled or partially labeled data, reducing the reliance on large labeled datasets.
Explainable AI: Developing models that can provide insights into their decision-making process, increasing trust and transparency.

Addressing Ethical Considerations:
Privacy: Ensuring the privacy and security of personal data collected by HAR systems.
Bias and Fairness: Mitigating biases in data and models to ensure fair and equitable treatment of all individuals.
Social Impact: Considering the potential social and ethical implications of widespread deployment of HAR technologies.

Real-World Deployment and Integration:
Developing robust and scalable systems that can be deployed and integrated into real-world applications.
Addressing the challenges of real-time performance, power consumption, and resource constraints.
Ensuring user acceptance and trust through user-centered design and clear communication.

XI. CONCLUSIONS

Human Activity Recognition (HAR) has emerged as a pivotal area of research with the potential to revolutionize various fields, including healthcare, surveillance, human-computer interaction, and robotics. While significant progress has been made, several challenges remain.

Data Acquisition: Overcoming challenges related to lighting variations, viewpoint limitations, occlusion, and sensor noise is crucial for robust and reliable HAR.
Data Representation: Effective representation of complex human actions, handling intra-class variability, and addressing the high dimensionality of data remain significant hurdles.
Computational Efficiency: Developing efficient algorithms and optimized implementations for real-time processing and resource-constrained environments is essential for practical applications.
Data Limitations: The availability of large, diverse, and unbiased datasets is crucial for training robust and generalizable models.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Despite these challenges, the future of HAR holds immense promise. Advancements in deep learning, multi-modal data fusion, and edge computing offer exciting opportunities to overcome existing limitations and unlock new applications.

Key future directions include:
Developing robust and generalizable models that can adapt to new environments and unseen activities.
Exploring unsupervised and semi-supervised learning techniques to reduce reliance on large labeled datasets.
Addressing ethical considerations such as privacy, bias, and fairness in the development and deployment of HAR systems.
Focusing on real-world deployment and integration by developing scalable and user-friendly systems.

REFERENCES

[1] Morshed, M. G., Sultana, T., Alam, A., & Lee, Y. K. (2015). Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities. IEEE Transactions on Multimedia, 17(4), 689-706.
[2] Zhou, Z., Shi, F., & Wu, W. (2015). Learning Spatial and Temporal Extents of Human Actions for Action Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(10), 2010-2023.
[3] Li, Y., Lan, C., Xing, J., Zeng, W., Yuan, C., & Liu, J. (2017). Online Human Action Detection using Joint Classification-Regression Recurrent Neural Networks. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2534-2543.
[4] Pinhanez, C., & Bobick, A. (1997). Human Action Detection Using PNF Propagation of Temporal Constraints. MIT Media Laboratory Perceptual Computing Section Technical Report No. 423.
[5] Barekatain, M., Martí, M., Shih, H.-F., Murray, S., Nakayama, K., Matsuo, Y., & Prendinger, H. (2017). Okutama-Action: An Aerial View Video Dataset for Concurrent Human Action Detection. Proceedings of IEEE International Conference on Computer Vision (ICCV), 3456-3464.
[6] Hu, Y., Cao, L., Lv, F., Yan, S., Gong, Y., & Huang, T. S. (2017). Action Detection in Complex Scenes with Spatial and Temporal Ambiguities. IEEE Transactions on Image Processing, 26(11), 5479-5490.
[7] Zhang, Y., Tokmakov, P., Hebert, M., & Schmid, C. (2017). A Structured Model For Action Detection. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011-2019.
[8] Ji, X., Cheng, J., Feng, W., & Tao, D. Skeleton embedded motion body partition for human action recognition using depth sequences. Signal Process. 2018, 143, 56–68.
[9] Lai, K., Bo, L., Ren, X., & Fox, D. A large-scale hierarchical multi-view rgb-d object dataset. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 1817–1824.
[10] Liu, M., & Liu, H. Depth context: A new descriptor for human activity recognition by using sole depth sequences. Neurocomputing 2016, 175, 747–758.
[11] Liu, M., Liu, H., & Chen, C. Robust 3D action recognition through sampling local appearances and global distributions. IEEE Trans. Multimed. 2017, 20, 1932–1947.
[12] Ji, X., Cheng, J., Feng, W., & Tao, D. Skeleton embedded motion body partition for human action recognition using depth sequences. Signal Process. 2018, 143, 56–68.
[13] Gowayyed, M.A., Torki, M., Hussein, M.E., & El-Saban, M. Histogram of oriented displacements (HOD): Describing trajectories of human joints for action recognition.
[14] M. D. Rodriguez, J. Ahmed, and M. Shah, “Action MACH: A spatiotemporal maximum average correlation height filter for action recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2008, pp. 1–8.
[15] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, “Learning realistic human actions from movies,” in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2008, pp. 1242–1249.
[16] H. Wang, A. Klaser, C. Schmid, and C.-L. Liu, “Action recognition by dense trajectories,” in Proc.
[17] Abdallah, Z.S., Gaber, M.M., Srinivasan, B., & Krishnaswamy, S. Activity recognition with evolving data streams: A review. ACM Comput. Surv. (CSUR) 2018, 51, 1–36.
[18] Herath, S., Harandi, M., & Porikli, F. Going deeper into action recognition: A survey. Image Vis. Comput. 2017, 60, 4–21.
[19] Jalal, A., Kim, Y.H., Kim, Y.J., Kamal, S., & Kim, D. Robust human activity recognition from depth video using spatiotemporal multi-fused features. Pattern Recognit. 2017, 61, 295–308.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[20] Yang, X., & Tian, Y. Super normal vector for human activity recognition with depth cameras. IEEE Trans. Pattern Anal. Mach. Intell. 2016, 39, 1028–1039.
[21] Xu, C., Govindarajan, L.N., & Cheng, L. Hand action detection from ego-centric depth sequences with error-correcting Hough transform. Pattern Recognit. 2017, 72, 494–503.
[22] Qi, J., Yang, P., Hanneghan, M., Tang, S., & Zhou, B. A hybrid hierarchical framework for gym physical activity recognition and measurement using wearable sensors. IEEE Internet Things J. 2018, 6, 1384–1393.
[23] Chen, C., Liu, K., & Kehtarnavaz, N. Real-time human action recognition based on depth motion maps. J. Real-Time Image Process. 2016, 12, 155–163.
[24] Nunez, J.C., Cabido, R., Pantrigo, J.J., Montemayor, A.S., & Velez, J.F. Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition. Pattern Recognit. 2018, 76, 80–94.
[25] Li, W., Zhang, Z., & Liu, Z. Action recognition based on a bag of 3d points. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13–18 June 2018; IEEE: Piscataway, NJ, USA, 2010; pp. 9–14.
[26] Bulbul, M.F., Jiang, Y., & Ma, J. Human action recognition based on dmms, hogs and contourlet transform. In Proceedings of the 2015 IEEE International Conference on Multimedia Big Data, Beijing, China, 20–22 April 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 389–394.
[27] Chen, C., Liu, M., Liu, H., Zhang, B., Han, J., & Kehtarnavaz, N. Multi-temporal depth motion maps-based local binary patterns for 3-D human action recognition. IEEE Access 2017, 5, 22590–22604.
[28] Zhang, B., Yang, Y., Chen, C., Yang, L., Han, J., & Shao, L. Action recognition using 3D histograms of texture and a multi-class boosting classifier. IEEE Trans. Image Process. 2017, 26, 4648–4660.
[29] Yang, X., Zhang, C., & Tian, Y. Recognizing actions using depth motion maps-based histograms of oriented gradients. In Proceedings of the 20th ACM International Conference on Multimedia, Nara, Japan, 29 October–2 November 2012; pp. 1057–1060.
[30] Oreifej, O., & Liu, Z. Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 716–723.