International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Smart Human Action Monitoring Using RGB and Motion Signals

Muskan Begum, department of Computer Science and Engineering, GNITC, 22-5K9, 22wj1a05k9@gniindia.org
MD. Saniya Anjum, department of Computer Science and Engineering, GNITC, 22-5J4, 22wj1a05j4@gniindia.org
MD Rahel Raza Hyder, department of Computer Science and Engineering, GNITC, 22-5J1, 22wj1a05j1@gniindia.org

---***---

Abstract - Human action recognition is a vital component of smart monitoring systems, with applications in healthcare, surveillance, and intelligent indoor environments. This project presents a smart human action monitoring system that relies exclusively on RGB images to detect and classify human activities in real time. The system processes image frames to extract essential features such as body posture, joint positions, and movement patterns. These features are then analyzed to recognize common human actions, including sitting, walking, running, drinking, and other daily activities. By leveraging advanced image processing and deep learning techniques, the system achieves high accuracy, robustness to varying lighting conditions, and efficiency suitable for real-time deployment. Experimental results demonstrate the system's ability to monitor human activity reliably, providing a practical solution for indoor action recognition and smart environment applications.

1.INTRODUCTION

Human Action Recognition (HAR) has become an essential area of research in computer vision due to its wide range of applications in healthcare, security surveillance, elderly monitoring, and intelligent indoor environments. The ability to automatically identify and classify human activities from visual data provides significant advantages in building smart systems that can support safety, automation, and human–computer interaction. For instance, monitoring the daily activities of elderly individuals can help detect abnormal behaviors, while in surveillance, action recognition enables the identification of suspicious activities in real time.

Traditional approaches to HAR often rely on wearable sensors, depth cameras, or motion capture systems to track body movements. While these methods can provide reliable information, they suffer from limitations such as high cost, intrusiveness, and dependency on specialized hardware. In contrast, RGB cameras offer a low-cost and non-intrusive solution that can be easily deployed in indoor environments. However, recognizing human activities from RGB images alone is challenging due to factors like varying lighting conditions, occlusions, and diverse human motion patterns.

Recent advancements in image processing and deep learning have significantly improved the performance of RGB-based HAR systems. By extracting key visual features such as body posture, joint positions, and temporal movement patterns, deep learning models can achieve high accuracy and robustness in action classification tasks. These improvements open the door to practical real-time monitoring solutions suitable for everyday environments.

In this work, we present a smart human action monitoring system that leverages RGB images to detect and classify human activities in real time. The proposed system is designed to recognize common daily actions such as sitting, walking, running, and drinking, with an emphasis on efficiency and robustness under different environmental conditions. By integrating advanced feature extraction techniques with deep learning models, our approach provides a reliable and practical framework for indoor action recognition, contributing to the development of intelligent and adaptive monitoring systems.

Key Words: Human Action Recognition, RGB Images, Deep Learning, Computer Vision, Activity Classification.

2. LITERATURE REVIEW

Title: A Hybrid Deep Learning-Based Approach for Human Activity Recognition Using

Wearable Sensors

Author: D. Sharma, A. Roy, S. P. Bag, P. K. Singh, and Y. Badr

Year: 2023

Description: Human Activity Recognition (HAR) is a branch of computer science that uses raw time-series data information from embedded smartphone sensors and wearable devices to infer human actions. It has aroused considerable interest in various smart home contexts, particularly for constantly monitoring human behavior in an ecologically friendly atmosphere for elderly people and rehabilitation. Data collection, feature extraction from noise and distortion, feature, and pre-processing and categorization are among the operating components of a typical HAR system. Extraction of feature and selection strategies have recently been developed using cutting-edge approaches and traditional machine learning classifiers. The majority of the solutions, on the other hand, rely on simple feature extraction algorithms that are unable to detect complex behaviors. Deep learning techniques are often utilized in different HAR approaches to recover features and classification swiftly because of the introduction and development of vast computing resources. The vast majority of solutions, on the other hand, depend on simplistic feature extraction algorithms incapable of recognizing complicated behaviors. Due to advancements in high computational capabilities, deep learning algorithms are now often utilized in HAR methods to efficiently extract meaningful features which can successfully categorize sensor data. In this chapter, we present a hybrid deep learning-based classification model comprising of Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM), which is named CNN-LSTM. The proposed hybrid deep learning model has been tested over three benchmark HAR datasets: MHEALTH, OPPORTUNITY, and HARTH. On the aforementioned datasets, the proposed hybrid model obtained 99.07%, 95.2%, and 94.68% classification accuracies, respectively, which is quite impressive

Title: Video-based human activity recognition: theory, methods and applications

Author: T. F. N. Bukht, H. Rahman, M. Shaheen, A. Algarni, N. A. Almujally, and A. Jalal

Year: 2024.

Description: Video-based human activity recognition (HAR) is an important task in many fields,such as healthcare monitoring, video surveillance, and sports analysis. This review paper aims to give an in-depth look at the current state of the art in HAR from 2018 to 2024. This will include a discussion of the different methods and models used for extracting, representing, and classifying human actions from video, as well as the challenges and limits of this field. The paper will also discuss recent improvements and plans for making HAR systems more accurate and useful. Even though there has been a lot of progress, a few knowledge gaps still need to be filled to make recognition more accurate and efficient. The purpose of this review paper is to offer scholars and professionals an overview of the theory, methods, and applications of HAR in videos. Through a critical analysis of the extant literature, this paper seeks to identify prospective avenues for future research and contribute towards advancing HAR systems that are more precise and efficient. By showing the different ways that HAR can be used, the paper shows how important this field is in many different areas.

Title: A Real-Time Human Action Recognition Model for Assisted Living

Author: Y. Wang, C. Muntean, P. Pathak, and P. Stynes

Year: 2024

Description: Ensuring the safety and well-being of elderly and vulnerable populations in assisted living environments is a critical concern. Computer vision presents an innovative and powerful approach to predicting health risks through video monitoring, employing human action recognition (HAR) technology. However, real-time prediction of human actions with high performance and efficiency is a challenge. This research proposes a real-time human action recognition model that combines a deep learning model and a live video prediction and alert system, in order to predict falls, staggering and chest pain for residents in assisted living. Six thousand RGB video samples from the NTU RGB+D 60 dataset were selected to create a dataset with four classes: Falling, Staggering, Chest Pain, and Normal, with the Normal class comprising 40 daily activities. Transfer learning technique was applied to train four state-of-the-art HAR models on a GPU server, namely, UniFormerV2, TimeSformer, I3D, and SlowFast. Results of the four models are presented in this paper based on class-wise and macro performance metrics, inference efficiency, model complexity and computational costs. TimeSformer is proposed for developing the real-time human action recognition model, leveraging its leading macro F1 score (95.33%), recall (95.49%), and precision (95.19%) along with significantly higher inference throughput compared to the others. This research provides insights to enhance safety and health of the elderly and people with chronic illnesses in assisted living environments, fostering sustainable care, smarter communities and industry innovation.

Title: Video-based Human Action Recognition using Deep Learning: A Review

Author: H. H. Pham, L. Khoudour, A. Crouzil, P. Zegers, and S. A. Velastin

Year: 2022

Description: Human action recognition is an important application domain in computer vision. Its primary aim is to accurately describe human actions and their interactions from a previously unseen data sequence acquired by sensors. The ability to recognize, understand, and predict complex human actions enables the construction of many important applications such as intelligent surveillance systems, human-computer interfaces, health care, security, and military applications. In recent years, deep learning has been given particular attention by the computer vision community. This paper presents an overview of the current state-of-the-art in action recognition using video analysis with deep learning techniques. We present the most important deep learning models for recognizing human actions, and analyze them to provide the current progress of deep learning algorithms applied to solve human action recognition problems in realistic videos highlighting their advantages and disadvantages. Based on the quantitative analysis using recognition accuracies reported in the literature, our study identifies state-of-the art deep architectures in action recognition and then provides current trends and open problems for future works in this field

.Title: A Survey on Human Action Recognition

Author: Zhou Shuchang

Year: 2022

Description: Human Action Recognition (HAR), one of the most important tasks in computer vision, has developed rapidly in the past decade and has a wide range of applications in health monitoring, intelligent surveillance, virtual reality, human computer interaction and so on. Human actions can be represented by a wide variety of modalities, such as RGB-D cameras, audio, inertial sensors,etc. Consequently, in addition to the mainstream single modality based HAR approaches, more and more research is devoted to the multimodal domain due to the complementary properties between multimodal data. In this paper, we present a survey of HAR methods in recent years according to the different input modalities. Meanwhile, considering that most of the recent surveys on HAR focus on the third perspective, while this survey aims to provide a more comprehensive introduction to HAR novices and researchers, we therefore also investigate the actions recognition methods from the first perspective in recent years. Finally, we give a brief introduction about the benchmark HAR datasets and show the performance comparison of different methods on these datasets.

3. RELATED WORK

This project focuses on the development of a smart human action recognition system that utilizes RGB images to detect and classify human activities in real time. The primary goal is to provide an efficient, cost-effective, and non-intrusive monitoring solution for applications in healthcare, surveillance, and intelligent indoor environments. The system targets the recognition of common daily actions such as sitting, walking, running, and drinking, which are crucial for activity tracking and safety monitoring in everyday life. The proposed approach leverages YOLOv8n, a state-of-the-art deep learning model, to automatically extract spatial and temporal features from image frames without the need for manual feature engineering. By combining real-time object detection and classification into a single pipeline, the system can accurately identify human postures and movement patterns. This integration ensures faster processing, robustness to environmental variations, and the ability to handle challenges such as low lighting, occlusions, and background clutter. To validate the system, experiments are conducted in indoor environments, where performance is evaluated in terms of accuracy, efficiency, and adaptability. The results demonstrate that the YOLOv8n-based system significantly outperforms traditional machine learning methods like SVM and Random Forest, offering high reliability for real-world deployment. Ultimately, this project presents a practical and scalable solution for human action monitoring, paving the way for advancements in smart healthcare systems, intelligent surveillance, and human–computer interaction.

3.1 METHODOLOGIES

3.1.1MODULES NAME:

Modules Name:

➢ Data Acquisition Module

➢ Preprocessing Module

➢ Feature Extraction Module
➢ Action Detection & Classification Module
➢ Real-Time Monitoring Module
➢ Result Visualization Module
➢ Performance Evaluation Module

3.1.2 MODULES EXPLANATION:

1) Data Acquisition Module:
This module is responsible for collecting input data in the form of RGB images or video streams from cameras or datasets. It serves as the foundation of the system by providing continuous visual input for analysis. The module ensures that the captured data maintains sufficient quality for further processing, even under varying environmental conditions. By supporting both live camera feeds and stored video files, the system can be applied in real-time monitoring as well as offline testing.

2) Preprocessing Module:
Before feeding the captured frames into the model, preprocessing is essential to improve recognition accuracy. This module handles tasks such as resizing frames to match the YOLOv8n input dimensions, normalizing pixel values, and enhancing image quality to reduce the impact of lighting variations. Noise removal and frame stabilization may also be applied to ensure consistent input. By standardizing the data, the preprocessing module ensures that the system remains robust across diverse environments and conditions.

3) Feature Extraction Module:
In traditional systems, feature extraction requires manual design, but in this project, the YOLOv8n model automatically extracts meaningful features from RGB frames. This module focuses on identifying spatial features like body posture and joint positions, along with temporal movement patterns. The deep learning backbone of YOLOv8n captures high-level representations of human activity without explicit hand-crafted feature engineering. These learned features form the basis for accurate action detection and classification.

4)Action Detection & Classification Module:
This module is the core of the proposed system. Using YOLOv8n, it detects humans within the input frame and classifies their actions into predefined categories such as sitting, walking, running, and drinking. The model processes frames in real time and assigns bounding boxes, labels, and confidence scores to each detected action. This integration of detection and classification ensures efficiency and simplifies the overall pipeline, making the system highly suitable for real-time applications.

5) Real-Time Monitoring Module:
The real-time monitoring module ensures continuous tracking of human actions for applications like healthcare monitoring, surveillance, and smart home automation. It processes the video feed frame by frame, updating action recognition results instantly. This module is designed to handle multiple individuals simultaneously and adapt to dynamic indoor environments. Its ability to deliver immediate feedback makes it highly practical for safety-critical scenarios, such as fall detection in elderly care or anomaly detection in security surveillance.

6) Result Visualization Module:
To make the system user-friendly and interpretable, the result visualization module overlays detection results directly on the video stream. It displays bounding boxes around individuals, labels for recognized actions, and confidence scores that indicate the reliability of predictions.This visual feedback not only aids in monitoring but also provides transparency in the system's decision-making process. The module can be extended to generate logs, reports, or alerts depending on the application

7) Performance Evaluation Module:
To make the system user-friendly and interpretable, the result visualization module overlays detection results directly on the video stream. It displays bounding boxes around individuals, labels for recognized actions, and confidence scores that indicate the reliability of predictions. This visual feedback not only aids in monitoring but also provides transparency in the system's decision-making process. The module can be extended to generate logs, reports, or alerts depending on the application.

4. PROPOSED METHODOLOGY

4.1 TECHNIQUE USED OR ALGORITHM USED

4.1.1 EXISTING TECHNIQUE:

➢ Support Vector Machine (SVM) and Random Forest
The existing technique for human action recognition in this project is based on traditional machine learning classifiers such as Support Vector Machine (SVM) and Random Forest (RF). These methods rely on manually extracted features from RGB image frames, such as body posture descriptors, motion vectors, or handcrafted appearance-based features. Once

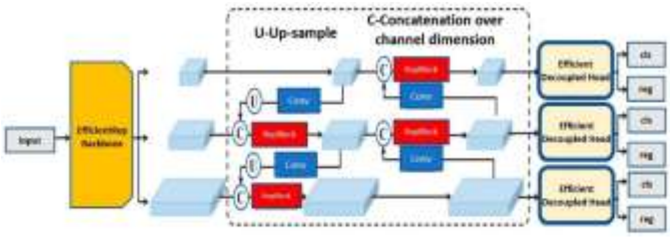
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

extracted, these features are fed into the classifiers to categorize actions like sitting, walking, or running. SVM operates by finding the optimal hyperplane that separates action classes in a high dimensional feature space. It is effective in handling small to medium-sized datasets and can achieve reasonable accuracy when the features are well-defined. Random Forest, on the other hand, is an ensemble-based approach that combines multiple decision trees to make robust predictions. It can handle noisy data better than individual classifiers and provides feature importance rankings, which helps in understanding the contribution of different features toward classification. Although these approaches have been widely used in earlier human action recognition studies, they have notable limitations. Their performance is highly dependent on the quality of handcrafted features, making them less effective in complex and dynamic real-world scenarios. They also struggle with temporal dependencies in actions, variations in lighting, occlusions, and cluttered backgrounds. Additionally, these models are not optimized for real-time monitoring, which restricts their practical deployment in intelligent systems.

4.1.2PROPOSED TECHNIQUE USED OR ALGORITHM USED:

➢ YOLOv8n: The proposed technique employs YOLOv8n (You Only Look Once, version 8 – nano variant), a state-of-the-art deep learning model, to perform real-time human action recognition using RGB images. Unlike traditional methods that rely on handcrafted features and separate classification models, YOLOv8n integrates feature extraction, detection, and classification into a single end to-end framework. This allows the system to directly process input frames, identify humans, and classify their actions simultaneously with high efficiency and accuracy. YOLOv8n is designed with a lightweight yet powerful architecture that makes it suitable for real time applications, even in resource-constrained environments. The model automatically learns spatial and temporal features such as body posture, movement patterns, and joint alignments from the input frames. By leveraging convolutional neural networks (CNNs) and advanced feature representation, it overcomes the limitations of handcrafted features used in existing systems. The nano variant (YOLOv8n) ensures reduced computational overhead while maintaining competitive accuracy, enabling deployment on edge devices like surveillance cameras or embedded systems. Furthermore, the proposed technique enhances robustness against challenges like varying lighting conditions, background clutter, and partial occlusions. The system outputs bounding boxes, action labels, and confidence scores in real time, making it highly practical for applications in healthcare monitoring, surveillance, and smart indoor environments. By combining speed, accuracy, and adaptability, the YOLOv8n-based approach provides a scalable solution for human action recognition that significantly outperforms traditional machine learning techniques such as SVM and Random Forest.

5.SYSTEM ARCHITECTURE



System Architecture

The proposed Smart Human Action Monitoring system follows a deep learning–based architecture designed to detect and classify human activities from RGB images in real time. The architecture processes video frames through several stages to extract meaningful information and recognize human actions efficiently. Initially, RGB frames captured from the input source are provided to the feature extraction backbone network. In this system, the EfficientRep backbone is used to extract high-level spatial features from the input images. This backbone network processes the images through multiple convolutional layers to capture important visual information related to human posture and movement. After feature extraction, the processed feature maps are passed to the neck network. The neck component performs operations such as up-sampling and channel concatenation to combine multi scale feature representations. These operations help the model retain both detailed and contextual information, which is important for detecting objects and recognizing actions at different scales. The architecture also incorporates RepBlock modules and convolution layers that improve feature representation and enhance the detection performance of the system. These components enable the model to learn complex patterns associated with human motion and posture. Finally, the extracted features are forwarded to the efficient decoupled head of the network. This head performs two important tasks simultaneously. The classification branch determines the type of human activity being performed, while the regression branch predicts the bounding box coordinates that locate the detected person in the frame. By combining detection and classification within a single architecture, the system can perform accurate and efficient real-time human action recognition.

6. RESULTS AND DISCUSSION

This project implements applications using Python, and the Server process is maintained using the SOCKET & SERVERSOCKET, and the Design part is played by Cascading Style Sheet.

Home Page

The Smart Human Action Monitoring system was evaluated to analyze its capability to detect and classify human activities from RGB video frames. The deep learning model based on YOLOv8 successfully identified human subjects and recognized several activities in real-time video streams. During experimental testing, the system demonstrated effective recognition of actions such as sitting, walking, running, and drinking. The model was able to generate accurate bounding boxes around detected individuals and assign the appropriate action labels along with confidence scores. The performance of the system was particularly strong in indoor environments with stable lighting conditions. By utilizing deep learning techniques, the system automatically learned spatial and motion related features directly from RGB images. This eliminated the need for manually designed feature extraction methods. The results also showed that the system can process video frames continuously while maintaining efficient processing speed. This capability makes the proposed approach suitable for applications that require real-time monitoring. Overall, the evaluation indicates that the system provides reliable detection performance and can effectively monitor human activities using only RGB visual data.

SignUp page

Login Page

Prediction Page

Result Page

6.CONCLUSION

In this project, a smart human action recognition system has been developed using YOLOv8n to detect and classify human activities in real time based on RGB images. Unlike traditional machine learning techniques such as SVM and Random Forest, which rely on handcrafted features, the proposed system leverages deep learning to automatically extract meaningful spatial and temporal features from image frames. This enables accurate recognition of daily activities like sitting, walking, running, and drinking with high robustness against environmental challenges such as varying lighting, occlusions, and cluttered backgrounds. The system demonstrates significant improvements in terms of accuracy, efficiency, and adaptability compared to existing methods. Its lightweight architecture ensures real-time performance while maintaining competitive accuracy, making it suitable for deployment in practical scenarios such as healthcare monitoring, smart homes, and surveillance applications. Overall, the project presents a cost-effective, non-intrusive, and scalable solution for human action recognition. With future enhancements such as support for more complex activities, multimodal data integration, and edge deployment, the system holds great potential to evolve into a comprehensive intelligent monitoring framework capable of addressing real-world challenges effectively.

7. FUTURE SCOPE

In the future, the cyberbullying detection system can be extended to support multilingual and cross-platform analysis, enabling detection across diverse languages and social media networks. Integration of transformer-based architectures such as BERT or RoBERTa could further enhance contextual understanding and semantic accuracy. Incorporating audio and image-based bullying detection will make the system more comprehensive for multimedia content. Real-time API deployment can allow integration with live chat applications or comment sections. The use of explainable AI (XAI) can help interpret model decisions and improve transparency. Adaptive online learning mechanisms can help the system evolve with changing slang and language styles. Introducing graph-based user behavior analysis can identify repeated offenders or coordinated harassment patterns. A dashboard interface can be developed for administrators to monitor and manage detected cases. Future research may focus on reducing bias and improving fairness across demographic groups. Overall, these enhancements will strengthen the system's scalability, accuracy, and social impact.

REFERENCES

[1] M. Muneeb, H. Rustam, and A. Jalal, ''Automate appliances via gestures recognition for elderly living assistance,'' in Proc. 4th Int. Conf. Advancement Comput. Sci. (ICACS), Feb. 2023, pp. 1–6.

[2] I. A. Abro and A. Jalal, ''Multi-modal sensors fusion for fall detection and action recognition in indoor environment,'' in Proc. 3rd Int. Conf. Emerg. Trends Electr., Control, Telecommun. Eng. (ETECTE), Nov. 2024, pp. 1–6.

[3] A. Nadeem, A. Jalal, and K. Kim, ''Accurate physical activity recognition using multidimensional features and Markov model for smart health fitness,'' Symmetry, vol. 12, no. 11, p. 1766, Oct. 2020.

[4] S. Hafeez, A. Jalal, and S. Kamal, ''Multi-fusion sensors for action recognition based on discriminative motion cues and random forest,'' in Proc. Int. Conf. Commun. Technol. (ComTech), Sep. 2021, pp. 91–96.

[5] A. Jalal, A. Nadeem, and S. Bobasu, ''Human body parts estimation and detection for physical sports movements,'' in Proc. 2nd Int. Conf. Commun., Comput. Digit. Syst. (C-CODE), Mar. 2019, pp. 104–109.

[6] A. Jalal, Y. Kim, and D. Kim, ''Ridge body parts features for human pose estimation and recognition from RGB-D video data,'' in Proc. 5th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT), Jul. 2014, pp. 1–6.

[7] A. Nadeem, A. Jalal, and K. Kim, ''Human actions tracking and recognition based on body parts detection via artificial neural network,'' in Proc. 3rd Int. Conf. Advancements Comput. Sci. (ICACS), Feb. 2020, pp. 1–6.

[8] M. A. K. Quaid and A. Jalal, ''Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm,'' Multimedia Tools Appl., vol. 79, nos. 9–10, pp. 6061–6083, Dec. 2019.

[9] A. Jalal and M. Mahmood, ''Students' behavior mining in e-learning environment using cognitive processes with information technologies,'' Educ. Inf. Technol., vol. 24, no. 5, pp. 2797– 2821, Sep. 2019.

[10] A. Jalal, N. Sarif, J. T. Kim, and T.-S. Kim, ''Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring

services at smart home,'' Indoor Built Environ., vol. 22, no. 1, pp. 271–279, Feb. 2013.

[11] M. G. Morshed, T. Sultana, A. Alam, and Y.-K. Lee, ''Human action recognition: A taxonomy-based survey, updates, and opportunities,'' Sensors, vol. 23, no. 4, p. 2182, Feb. 2023, doi: 10.3390/s23042182.

[12] M. Pervaiz and A. Jalal, ''Artificial neural network for human object interaction system over aerial images,'' in Proc. 4th Int. Conf. Advancements Comput. Sci. (ICACS), Feb. 2023, pp. 1–

6.

[13] B. Ren, M. Liu, R. Ding, and H. Liu, ''A survey on 3D skeleton-based action recognition using learning method,'' Cyborg Bionic Syst., vol. 5, Jan. 2024, Art. no. 0100, doi:

10.34133/cbsystems.0100.

[14] Y. Jia, G. Chen, and H. Chi, ''Retinal fundus image super-resolution based on generative adversarial network guided with vascular structure prior,'' Sci. Rep., vol. 14, no. 1, p. 22786, Oct. 2024, doi: 10.1038/s41598-024- 74186-x.

[15] Z. Liu, Y. Gu, L. Yu, X. Yang, Z. Ma, J. Zhao, and Y. Gu, ''Locomotion control of cyborg insects by charge-balanced biphasic electrical stimulation,'' Cyborg Bionic Syst., vol. 5, Jan.

2024, Art. no. 0134, doi: 10.34133/cbsystems.0134.

[16] X. Zhao, T. Wang, Y. Li, B. Zhang, K. Liu, D. Liu, C. Wang, and H. Snoussi, ''Targetdriven visual navigation by using causal intervention,'' IEEE Trans. Intell. Vehicles, vol. 9, no. 1, pp. 1294–1304, Jan. 2024, doi: 10.1109/TIV.2023.3288810.

[17] M. Mahmood, A. Jalal, and K. Kim, ''WHITE STAG model: Wise human interaction tracking and estimation (WHITE) using spatio-temporal and angular-geometric (STAG) descriptors,'' Multimedia Tools Appl., vol. 79, nos. 11–12, pp. 6919–6950, Mar. 2020. [18] U. Azmat and A. Jalal, ''Smartphone inertial sensors for human locomotion activity recognition based on template matching and codebook generation,'' in Proc. Int. Conf. Commun. Technol. (ComTech), Sep. 2021, pp. 109–114.

[19] A. Jalal, M. Mahmood, and A. S. Hasan, ''Multi-features descriptors for human activity tracking and recognition in indoor-outdoor environments,'' in Proc. 16th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), Jan. 2019, pp. 371–376.

[20] A. Nadeem, A. Jalal, and K. Kim, ''Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy Markov model,'' Multimedia Tools Appl., vol. 80, no. 14, pp. 21465–21498, Jun. 2021, doi: 10.1007/s11042-021-10687-5.

[21] R. Jansi and R. Amutha, ''Sparse representation-based classification scheme for human activity recognition using smartphones,'' Multimedia Tools Appl., vol. 78, no. 8, pp. 11027– 11045, Sep. 2018, doi: 10.1007/s11042-018-6662-5.

[22] Y. Tian, X. Wang, L. Chen, and Z. Liu, ''Wearable sensor-based human activity recognition via two-layer diversity-enhanced multiclassifier recognition method,'' Sensors, vol. 19, no. 9, p. 2039, Apr. 2019, doi: 10.3390/s19092039.

[23] P. Kumar, S. Chauhan, and L. K. Awasthi, ''Human activity recognition (HAR) using deep learning: Review, methodologies, progress and future research directions,'' Arch. Comput. Methods Eng., vol. 31, no. 1, pp. 179–219, Aug. 2023, doi: 10.1007/s11831-023-09986-x. [24] H. Haresamudram, A. Beedu, V. Agrawal, P. L. Grady, and I. Essa, ''Masked reconstruction based self-supervision for human activity recognition,'' in Proc. 24th Annu. Int. Symp. Wearable Comput., Cancun, Mexico, 2020, pp. 45–49, doi: 10.1145/3410531.3414306. [25] A. Jordao, A. C. Nazare Jr., J. Sena, and W. R. Schwartz, ''Human activity recognition based on wearable sensor data: A standardization of the stateof-the-art,'' 2018, arXiv:1806.05226.

[26] M. Batool, A. Jalal, and K. Kim, ''Sensors technologies for human activity analysis based on SVM optimized by PSO algorithm,'' in Proc. Int. Conf. Appl. Eng. Math. (ICAEM), Taxila, Pakistan, Aug. 2019, pp. 145–150, doi: 10.1109/ICAEM.2019.8853770.

[27] N. Archana and K. Hareesh, ''Real-time human activity recognition using ResNet and 3D convolutional neural networks,'' in Proc. 2nd Int. Conf. Adv. Comput., Commun., Embedded

Secure Syst. (ACCESS), Sep. 2021, pp. 173–177, doi: 10.1109/ACCESS51619.2021. 9563316. [28] J. Lee and B. Ahn, ''Real-time human action recognition with a low-cost RGB camera and mobile robot platform,'' Sensors, vol. 20, no. 10, p. 2886, May 2020, doi: 10.3390/s20102886. [29] N. Crasto, P. Weinzaepfel, K. Alahari, and C. Schmid, ''MARS: Motionaugmented RGB stream for action recognition,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 7874–7883, doi: 10.1109/CVPR.2019.00807.

[30] I. M. Nasir, M. Raza, J. H. Shah, M. Attique Khan, and A. Rehman, ''Human action recognition using machine learning in uncontrolled environment,'' in Proc. 1st Int. Conf. Artif.

Intell. Data Analytics (CAIDA), Apr. 2021, pp. 182–187, doi: 10.1109/CAIDA51941.2021. 9425202.

I. A. Abro et al.: Advanced Biosignal-RGB Fusion With Adaptive Neurofuzzy Classification

[31] A. Jalal, I. Akhtar, and K. Kim, ''Human posture estimation and sustainable events classification via pseudo-2D stick model and K-ary tree hashing,'' Sustainability, vol. 12, no. 23, p. 9814, Nov. 2020, doi: 10.3390/su12239814.

[32] G. L. Goh, G. D. Goh, J. W. Pan, P. S. P. Teng, and P. W. Kong, ''Automated service height fault detection using computer vision and machine learning for badminton matches,'' Sensors, vol. 23, no. 24, p. 9759, Dec. 2023, doi: 10.3390/s23249759.

[33] B. Debnath, M. O'Brien, M. Yamaguchi, and A. Behera, ''A review of computer visionbased approaches for physical rehabilitation and assessment,'' Multimedia Syst., vol. 28, no. 1, pp. 209–239, Feb. 2022, doi: 10.1007/s00530-021-00815-4.

[34] Y. Hou, J. Li, S. Liao, and N. Xue, ''Research advanced in human pose estimation based on deep learning,'' Highlights Sci., Eng. Technol., vol. 119, pp. 444–453, Dec. 2024, doi:

10.54097/f4948z63.

[35] S. U. Rehman, A. U. Yasin, E. Ul Haq, M. Ali, J. Kim, and A. Mehmood, ''Enhancing human activity recognition through integrated multimodal analysis: A focus on RGB imaging, skeletal tracking, and pose estimation,'' Sensors, vol. 24, no. 14, p. 4646, Jul. 2024, doi:

10.3390/s24144646.

[36] R. B. Neupane, K. Li, and T. F. Boka, ''A survey on deep 3D human pose estimation,''

Artif. Intell. Rev., vol. 58, no. 1, Nov. 2024, Art. no. 24, doi: 10.1007/s10462-024-11019-3. [37] P. Gupta, ''Review on human pose estimation using AI fitness tracker,'' Indian Sci. J. Res. Eng. Manag., vol. 8, no. 4, pp. 1–5, Apr. 2024, doi: 10.55041/ijsrem29924.

[38] G. Di Gennaro, A. Buonanno, M. Baldi, E. Capoluongo, and F. Palmieri, ''Vision-based human activity recognition methods using pose estimation,'' in Applications of Artificial

Intelligence and Neural Systems To Data Science. Cham, Switzerland: Springer, 2023, pp. 115– 125, doi: 10.1007/978-981-99-3592-5_11.

[39] C. M. Ranieri, S. MacLeod, M. Dragone, P. A. Vargas, and R. A. F. Romero, ''Activity recognition for ambient assisted living with videos, inertial units and ambient sensors,'' Sensors, vol. 21, no. 3, p. 768, Jan. 2021, doi: 10.3390/s21030768.

[40] H. Basak, R. Kundu, P. K. Singh, M. F. Ijaz, M. Woźniak, and R. Sarkar, ''A union of deep learning and swarm-based optimization for 3D human action recognition,'' Sci. Rep., vol. 12, no. 1, Mar. 2022, Art. no 5494, doi: 10.1038/s41598-022-09293-8.

[41] M. M. Islam and T. Iqbal, ''Multi-GAT: A graphical attention-based hierarchical multimodal representation learning approach for human activity recognition,'' IEEE Robot. Autom. Lett., vol. 6, no. 2, pp. 1729–1736, Apr. 2021, doi: 10.1109/LRA.2021.3059624.

[42] C. Chen, R. Jafari, and N. Kehtarnavaz, ''UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor,'' in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2015, pp. 168–172, doi: 10.1109/ICIP.2015.7350781.

[43] L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, and L. Van Gool, ''Temporal segment networks for action recognition in videos,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 11, pp. 2740–2755, Nov. 2019, doi: 10.1109/TPAMI.2018.2868668.

[44] S. K. Yadav, K. Tiwari, H. M. Pandey, and S. A. Akbar, ''A review of multimodal human activity recognition with special emphasis on classification,

applications, challenges and future directions,'' Knowl.-Based Syst., vol. 223, Jul. 2021, Art. no. 106970, doi:

10.1016/j.knosys.2021.106970.

[45] Q. Wu, Y. Li, H. Ouyang, and Y. Xuan, ''A multimodal fusion human activity recognition method based on CNN and gramian angular field,'' Proc. SPIE, vol. 12791, Oct. 2023, Art. no.

127911I, doi: 10.1117/12.3004953.

[46] M. Ehatisham-Ul-Haq, A. Javed, M. A. Azam, H. M. A. Malik, A. Irtaza, I. H. Lee, and M. T. Mahmood, ''Robust human activity recognition using multimodal feature-level fusion,'' IEEE Access, vol. 7, pp. 60736–60751, 2019, doi: 10.1109/ACCESS.2019.2913393.

[47] M. Amsaprabhaa, ''Hybrid optimized multimodal spatiotemporal feature fusion for visionbased sports activity recognition,'' J. Intell. Fuzzy Syst., vol. 46, no. 1, pp. 1481–1501, Jan. 2024, doi: 10.3233/jifs-233498.

[48] M. Conway, B. Reily, and C. Reardon, ''Learned sensor fusion for robust human activity recognition in challenging environments,'' in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.

(IROS), Oct. 2024, pp. 11537–11543, doi: 10.1109/iros58592.2024.10802830.

[49] A. Alazeb, U. Azmat, N. Al Mudawi, A. Alshahrani, S. S. Alotaibi, N. A. Almujally, and A. Jalal, ''Intelligent localization and deep human activity recognition through IoT devices,'' Sensors, vol. 23, no. 17, p. 7363, Aug. 2023, doi: 10.3390/s23177363.

[50] A. Arif and A. Jalal, ''Automated body parts estimation and detection using salient maps and Gaussian matrix model,'' in Proc. Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), Jan.

2021, pp. 667–672.

[51] A. R. Shahzad and A. Jalal, ''A smart surveillance system for pedestrian tracking and counting using template matching,'' in Proc. Int. Conf. Robot. Autom. Ind. (ICRAI), Oct. 2021, pp. 1–6.

[52] H. Ansar and A. Jalal, ''Robust hand gesture tracking and recognition for healthcare via recurrent neural network,'' in Proc. 20th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), Aug. 2023, pp. 148–153.

[53] K. Saleha and J. Ahmad, ''A novel human interaction recognition via composite features and Max entropy classifier,'' in Proc. ICET, 2024, pp. 1–6.

[54] L. Zahoor and A. Jalal, ''Drone-based human surveillance using YOLOv5 and multifeatures,'' in Proc. Int. Conf. Frontiers Inf. Technol. (FIT), Dec. 2024, pp. 1–6.

[55] L. Zahoor and A. Jalal, ''Remote sensing based human monitoring and recognition over public surveillance,'' in Proc. 3rd Int. Conf. Emerg. Trends Electr., Control, Telecommun. Eng. (ETECTE), Nov. 2024, pp. 1–6.

[56] M. Pervaiz, A. Jalal, and K. Kim, ''Hybrid algorithm for multi people counting and tracking for smart surveillance,'' in Proc. Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), Islamabad, Pakistan, Jan. 2021, pp. 530–535, doi: 10.1109/IBCAST51254.2021.9393171.

[57] A. Jalal, M. A. K. Quaid, S. B. U. D. Tahir, and K. Kim, ''A study of accelerometer and gyroscope measurements in physical life-log activities detection systems,'' Sensors, vol. 20, no. 22, p. 6670, Nov. 2020, doi: 10.3390/s20226670.

[58] M. Waheed, A. Jalal, M. Alarfaj, Y. Y. Ghadi, T. A. Shloul, S. Kamal, and D.-S. Kim, ''An LSTM-based approach for understanding human interactions using hybrid feature descriptors over depth sensors,'' IEEE Access, vol. 9, pp. 167434–167446, 2021, doi: 10.1109/ACCESS.2021.3130613.

[59] J.-T. Hsieh, M.-L. Chiang, C.-Y. Fang, and S.-W. Chen, ''Online human action recognition using deep learning for indoor smart mobile robots,'' in Proc. Int. Conf. Comput., Commun., Intell. Syst. (ICCCIS), Greater Noida, India, Feb. 2021, pp. 425–433, doi:

10.1109/ICCCIS51004.2021.9397242.

[60] U. Azmat, S. S. Alotaibi, N. A. Mudawi, B. I. Alabduallah, M. Alonazi, A. Jalal, and J. Park, ''An elliptical modeling supported system for human action deep recognition over aerial surveillance,'' IEEE Access, vol. 11, pp. 75671–75685, 2023, doi:

10.1109/ACCESS.2023.3266774.

[61] M. Batool, S. S. Alotaibi, M. H. Alatiyyah, K. Alnowaiser, H. Aljuaid, A. Jalal, and J. Park,

‘‘Depth sensors-based action recognition using a modified K-Ary entropy classifier,’’ IEEE Access, vol. 11, pp. 58578–58595, 2023, doi: 10.1109/ACCESS.2023.3260403. [62] A. Jalal, M. Batool, and K. Kim, ‘‘Stochastic recognition of physical activity and healthcare using tri-axial inertial wearable sensors,’’ Appl. Sci., vol. 10, no. 20, p. 7122, Oct. 2020, doi: 10.3390/app10207122.

[63] M. Batool, A. Jalal, and K. Kim, ‘‘Telemonitoring of daily activity using accelerometer and gyroscope in smart home environments,’’ J. Electr. Eng. Technol., vol. 15, no. 6, pp. 2801–2809, Nov. 2020, doi: 10.1007/s42835- 020-00554-y.

[64] T. Wang, J. Li, H.-N. Wu, C. Li, H. Snoussi, and Y. Wu, ‘‘ResLNet: Deep residual LSTM network with longer input for action recognition,’’ Frontiers Comput. Sci., vol. 16, no. 6, Dec. 2022, Art. no. 166334, doi: 10.1007/s11704-021-0236-9.

[65] M. Javeed, M. Abdelhaq, A. Algarni, and A. Jalal, ‘‘Biosensor-based multimodal deep human locomotion decoding via Internet of Healthcare Things,’’ Micromachines, vol. 14, no. 12,

p. 2204, Dec. 2023, doi: 10.3390/mi14122204.

[66] A. Jalal, M. Batool, and K. Kim, ‘‘Sustainable wearable system: Human behavior modeling for life-logging activities using K-ary tree hashing classifier,’’ Sustainability, vol. 12, no. 24, p. 10324, Dec. 2020, doi: 10.3390/su122410324.

[67] Y. Wu, Z. Jia, Y. Ming, J. Sun, and L. Cao, ‘‘Human behavior recognition based on 3D features and hidden Markov models,’’ in Proc. Signal, Image Video Process. (SIViP), vol. 10, Feb. 2015, pp. 495–502, doi: 10.1007/s11760-015-0756-6.

[68] A. Jalal, N. Khalid, and K. Kim, ‘‘Automatic recognition of human interaction via hybrid descriptors and maximum entropy Markov model using depth sensors,’’ Entropy, vol. 22, no. 8, p. 817, Jul. 2020.

[69] S. Majumder and N. Kehtarnavaz, ‘‘Vision and inertial sensing fusion for human action recognition: A review,’’ IEEE Sensors J., vol. 21, no. 3, pp. 2454–2467, Feb. 2021, doi:

10.1109/JSEN.2020.3022326.

[70] G. Xiao, Z. Xiao, P. Zhou, X. Jia, N. Wang, D. Zhao, and H. Wei, ‘‘PPP based on factor graph optimization,’’ Meas. Sci. Technol., vol. 35, no. 11, Nov. 2024, Art. no. 116307, doi:

10.1088/1361-6501/ad6680.

[71] Q. Sun, H. Wang, W. Liu, J. Zou, F. Ye, and Y. Li, ‘‘An improved stereo visual-inertial SLAM algorithm based on point-and-line features for subterranean environments,’’ IEEE Trans. Veh. Technol., vol. 74, no. 3, pp. 3925–3940, Mar. 2025, doi: 10.1109/TVT.2024.3492388. [72] T. Long, X. Song, B. Han, Y. Suo, and L. Jia, ‘‘In situ magnetic field compensation method for optically pumped magnetometers under three-axis nonorthogonality,’’ IEEE Trans. Instrum.

Meas., vol. 73, pp. 1–12, 2024, doi: 10.1109/TIM.2023.3331425.

[73] Y. He, M. Bao, Y. Chen, H. Ye, J. Fan, and G. Shi, ‘‘Accuracy characterization of Shack– Hartmann sensor with residual error removal in spherical wavefront calibration,’’ Light, Adv.

Manuf., vol. 4, no. 4, p. 1, 2023, doi: 10.37188/lam.2023.036.

[74] C. Li, H. Wang, Y. Wen, R. Yin, X. Zeng, and K. Li, ‘‘GenoM7GNet: An efficient N7 methylguanosine site prediction approach based on a nucleotide language model,’’ IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 21, no. 6, pp. 2258–2268, Nov. 2024, doi: 10.1109/TCBB.2024.3459870.

[75] G. Cai, X. Zheng, J. Guo, and W. Gao, ‘‘Real-time identification of borehole rescue environment situation in underground disaster areas based on multi-source heterogeneous data fusion,’’ Saf. Sci., vol. 181, Jan. 2025, Art. no. 106690, doi: 10.1016/j.ssci.2024.106690.