JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Video-Based Student Activity Recognition Using Deep Learning: A Comprehensive Review

Rakesh Kancharla[1], Srivalli Pinninti[2], Sri Harika Mani Doddaboina[3], Surya Venkata Bhavani Koujula[4], Lakshmi Bhavan Donga[5], M Parthipan[6]
Department of Computer Science and Engineering,Sasi Institute Of Technology & Engineering

Abstract — ***In an era of quickly growing educational frameworks and digital change, student activity recognition has become a vital area. Activity identification systems are a vast subject of research and development, presently with a focus on advanced machine learning, deep learning algorithms, and reducing the costs of monitoring. The research largely focuses on the applications of activity recognition systems and surveys. This complete article covers student activity recognition, covering its methodology, uses, and consequences in the area of education. The applications are categorized according to the methodology utilized for identifying human activity, namely as based on visual, non-visual, and multimodal sensor technologies. This presents an overview of different applications and explores the merits and limits of each strategy. Additionally, we present public datasets that are created for the assessment of such identification algorithms. Further The study finishes with a comparison of the available strategies that, when applied to real-world settings, enable the development of research topics for future approaches.***

Keywords — Student Activity Recognition (SVR), Human Activity Classification (HAC), Deep Learning Algorithms, Machine Learning Algorithms.

I. INTRODUCTION

Human activity recognition, an automatic identification of actions done by people from captured video, has been considered to be one of the foremost significant challenges in computer perception since the early 1990s. Human Activity Recognition (HAR) employs modern computer algorithms to analyse, categorize, and contextualize diverse activities and behaviors done by people.As a field that encompasses a diverse range of activities, from everyday movements to complex gestures, HAR has experienced significant growth. In recent years the field of activity recognition becomes an integral part of many social life systems that including smart surveillance quality, healthcare, security, and education.

One of the most promising applications of HAR lies in the field of education. As we seek to transform traditional classrooms into interactive and data-rich learning environments, the principles and techniques of HAR become instrumental. The recognition of student activities within the classroom, such as participation, attentiveness, and engagement, can profoundly impact pedagogical strategies and educational outcomes. By extending the capabilities of HAR, we can pave the way for Student Activity Recognition (SAR). SAR not only encompasses the physical activities of students but also their cognitive and behavioural engagement. Leveraging computer vision, sensor data, and machine learning, SAR has the potential to revolutionize the educational landscape by offering educators real-time insights into classroom dynamics and the learning experience. This invaluable data allows for personalized instruction, early intervention for struggling students, and evidence-based educational decisions.

As of today's educational, scene, the classroom experience is experiencing a significant transition, with technology playing a vital role in influencing teaching and learning. We want to provide instructors with useful insights regarding the habits and attentiveness of pupils. Ensuring reliable monitoring of their travels and engagement levels. The world of activity recognition has become a wide area of research and innovation, with a major focus on using futuristic machine learning and deep learning algorithms. Simultaneously, there is a determined attempt to cut the expenditures involved with supervision. Some famous algorithms in this area includes Convolution Neural Networks-Gated Recurrent Unit, Normalized mutual information-based feature selection (NMIFS), Fusion of Deep Learning Algorithms, all contributing to the evolution of activity recognition systems.

II. RELATED WORK

Activity Recognition of Human is contingent on deep-temporal learning employing convolutional neural networks (CNNs), as explained in [1]. The recommended approach consists of three crucial components are deep feature extraction, feature selection, and a bidirectional gated recurrent unit (Bi-GRU) for processing sequential data. The idea behind this method is to take inborn knowledge about human activity from pre-trained CNNs, identify the prominent traits among firmly constructed visualizations, and supply them to acquire time-related dynamics. The authors deploy bidirectional gated recurrent units to mimic sequential data and apply a random forest approach to decrease processing complexity. The recommended approach is evaluated on three baseline datasets: YouTube11, HMDB51, along with UCF101. These YouTube11 datasets comprise 1160 sports videos sorted into 11 groups, while the HMDB51 and UCF101 datasets

have 51 and 101 action categories, respectively. The recommended approach gives great accuracy in recognizing human activities and may be applied to numerous applications outside of human activity identification.

[2] The paper explores human activity identification based on still-based image data. The authors employ transfer learning and attention modules to enhance the accuracy of human activity identification from photos on these datasets. The plan of action incorporates four initially trained convolutional neural networks: InceptionV3, Xception, InceptionResnetV2, and EfficientNetB7, each of which is utilized to extract features from a different channel of the input picture. These traits are then concatenated using an attention approach to form a final feature vector, which is put into a completely interconnected layer for classification. In the future, the scientists suggest that there is a chance of developing a means to extract the skeleton joint locations of human activities using photographs in order to reach an even higher identification rate.

[3][8] These two projects work on radar motion data for human activity identification (HAR). [3] The research proposes a new approach for the classifying and finding of human activities depends on radar echoes. The proposed approach incorporates a complex-value convolutional neural network (CV-CNN) structure, which undergoes training using time-frequency (TF) representations of radar echoes collected from capturing motion records of human actions. The study also includes a complete investigation of the motion characteristics of individuals and the micro-Doppler effects induced by varied human activities, which is applied to develop a human activity sample database. The recommended approach obtains a precise classification rate of 99.81% on the test set and is resilient to noise and the amount of training samples. [8] This study proposes a new approach to handle the challenge of using millimeter-wave 4D imaging radars to categorize movements at undesirable aspect angles and static positions in between continuous sequences of activities. The suggested method includes a pipeline for hierarchical processing and classification that fully utilizes all the data from the radar. The pipeline uses the complementing point cloud and spectrogram data formats, and an experimental dataset with six activities completed by eight persons is used to demonstrate the pipeline's efficacy. The findings show that the recommended pipeline operates more effectively than numerous baseline approaches in the literature.

[4] The recommended approach leverages a deep learning CNN model to classify the input data and semantic labels, which reduces the necessity for explicit feature extraction and representation techniques. The data employed in this investigation originates from Florence 3D Action (MICC, 2018) and was gathered using a Kinect camera. The authors deep learning models were able to recognize activities with an optimum accuracy of 94.08% after using Euclidean distance to extract motion features from the dataset., which is larger than the accuracy reached by earlier comparable research. The research also evaluates comparable studies on the issue and presents an appraisal of the recommended system. Overall, the presented approach reveals a viable way to accurately recognize human behaviors using deep learning, which has potential applications in diverse areas, including healthcare, sports, security, and entertainment.

In their proposed study, [5] and [6] analyse sensor data from many channels in order to identify human actions. [5] They suggested HAR in the present research, a multichannel CNN-GRU paradigm was applied. The model includes two GRU layers following feature fusion, three independent channels with convolutional kernels of varied sizes, and GAP and Batch Norm layers feeding the final classification layer. The goal of the model is to enhance the neural network's vision by extracting properties of various scales from the input samples. On a number of benchmark datasets, such as WISDM, UCI-HAR, and PAMAP2, the model has been shown to achieve greater accuracy than existing deep learning algorithms. [6] The recommended method uses mixup triplets to train deep neural networks under open-set assumptions. It is based on deep learning. In order to rule out unknown actions as belonging to unknown classes, the technique also adds a Mahalanobis distance-based criterion for each known activity. On three sensor-based HAR datasets, the determined methodology is analysed and contrasted with already present methods. The output show that, in the context of efficiency and F1 score, the suggested strategy operates more effectively than the alternatives. In order to improve the recommended system while making it more applicable to real-world situations, the paper also makes a number of recommendations for future research areas of study.

[7] This research study provides a unique framework for human activity identification in videos utilizing transformer neural networks. The recommended technique in this study is the ViT-ReT architecture, which contains two transformer neural networks: the Vision Transformer (ViT) and the Recurrent Transformer (ReT). The ViT is used to extract spatial attributes from video frames, while the ReT is used to describe the temporal links between these features. The recommended ViT-ReT architecture gives a large speedup over traditional CNN and RNN models while maintaining equal accuracy. The authors prove the applicability of their technique on several publicly available datasets and show that it exceeds state-of-the-art approaches. The authors also highlight probable future research areas for extending their results. Overall, this work gives a feasible strategy for enhancing the speed and scalability of activity.

[9] Advanced Learning-Based Human Action Recognition in Drone Video" is an action recognition technique for individuals in aerial video clips. The system employs deep learning approaches, mainly convolutional neural networks (CNN), to identify human activity in RGB videos and extract properties from the footage It contains techniques like body key-point extraction, quick shift separation, elliptical analysis based on EM-GMM, and feature optimization making use of the Naive Bayes feature optimizer.The system seeks to boost the accuracy and robustness of human action detection in various practical applications including video surveillance, robot-human cooperation, sports analysis, and recreational activities. The authors also mention possible future research options for expanding their work towards Incorporating new modalities, strengthening feature extraction, handling complex and changeable datasets, boosting recognition accuracy, researching real-time applications, and examining transfer learning.

[10] Inertial sensor-based HAR systems utilize wearable sensors attached to various parts of the human body to determine human actions. This article includes comprehensive details on Investigational studies in this area, including the use of CNN with DFSMN to represent persistent dependencies in time sequences and the

integration of local features and scale-invariants with behavioural relationships using CNN and LSTM. However, additional study is essential to examine the effectiveness of these models in real-world conditions and compare them against other modern methods. For the categorization, we have applied technologies like pre-processing approaches, background removal techniques, etc. The suggested technique has been evaluated on three standard datasets, including the UP Fall data set, the University of Rzeszow Fall data set, and the SisFall data set. In the future, the authors propose to examine new characteristics and modelling methodologies associated with human activity recognition (HAR), leveraging deep learning techniques and difficult datasets. They also aim to boost the efficiency and effectiveness of the recommended hybrid descriptors for human activity detection.

[11] The article focuses on fall detection and activity identification using human skeletal characteristics collected from video footage. The basic premise is that employing articulated bodies (skeletons) generated from the video may considerably increase the performance of fall detection and activity identification compared to using other sensor modalities. The method integrates human skeletal identification using AlphaPose, a multi-person posture estimator, and the usage of four model classifiers (RF, SVM, MLP, and KNN) both fall detection as well as activity classification. The technique is studied using the UP-FALL dataset and displays greater performance compared to other strategies tested on the same dataset. The recommended method has the benefit of identifying human activity using a person's posture in a video clip, allowing for the identification of more than one person in the scene. Future work involves building an algorithm for multi-person identification and resolving confusion with the lying action. In future studies, the researchers want to build an algorithm that can handle multi-person recognition and solve the issue of lying.

[12] The recommended work of the research focuses on the design of a unique HDAR (Human Detection and Activity Recognition) technique for recognizing individuals and differentiating their actions from the captured video sequences above. The study highlights the application of EfficientDet-D7, a method of object detector, to boost the accuracy of human identification in videos involving changes such as elevation differences, illuminating changes, camera jitter, and variations in perspectives, object dimensions, and hues. The performance of three distinct deep learning models, namely MobileNetV2, DenseNet121, and ResNet18, is compared and assessed in the context of human activity identification. Future work may concentrate on expanding the capabilities of UAV systems for human identification, examining the possibilities of deep learning and AI in traffic surveillance, and further developing the use of drones in law enforcement operations.

[13] The recommended work of the study is a switching structured prediction technique for human activity recognition employing probabilistic graphical models. The approach is supposed to find both straightforward as well as challenging activities, and it requires dividing the issue into distinct parts and addressing smaller issues using the proposed probabilistic graphical models. The authors claim that this is the first time regime switching has been applied to HAR. The paper also evaluates related studies on the issue of probabilistic graphical models for activity detection and presents practical data to highlight the utility of the recommended method. This paper uses numerous strategies for the extraction of different parameters for HAR. Probabilistic Graphical Models (PGMs) for activity identification, Latent Structured Support Vector Machines (LSSVM) for learning parameters of graphical models, Maximum-entropy Markov model (MEMM) relies on hierarchical MEMM, Dual decomposition-based approach for inference on graphs with nontree and generic structures, Block-coordinate descent for maximizing the objective function

[14] The intended purpose of the work is to design a poorly supervised method for recognizing human activities from unconstrained videos employing intermediate spatial and a high-level prejudiced property. It believes that it has a single behaviour or activity per video clip to recognize and fails to utilize annotations offered by academics on existing datasets to train the model. The recommended approach is tested on four real-world existing datasets, including VIRAT 1.0 Ground and VIRAT 2.0 Ground. The paper compares the computational complexity of the suggested technique with different prominent algorithms. The recommended technique is assessed against their supervised approach for detecting human activities from unconstrained films utilizing mid-level contextual and high-level discriminative characteristics. The approach believes that there is just one action per video segment to detect and does not leverage annotations supplied by academics on existing datasets to train the model. The recommended approach is tested on four real-world existing datasets, including VIRAT 1.0 Ground and VIRAT 2.0 Ground. The paper compares the computational complexity of the suggested technique with different prominent algorithms. The suggested approach is examined against the baseline low-level features based on the popular spaial-time local features (HoG, HOF, and MBH) using an SVM model with a conventional radial basis function (RBF). The paper suggests some future directions for research, including exploring the use of deep learning techniques to learn more discriminative features from the data, investigating the use of multi-task learning to recognize multiple activities in a single video segment, and exploring the use of transfer learning to improve recognition performance on new datasets.

[15] This study proposes a unique feature selection approach for human activity detection systems named Normalized Mutual Information-based Feature Selection (NMIFS). The recommended methodology includes numerous feature extraction approaches to accurately differentiate human behaviors. The article addresses the processes used to assure that the selection process is not influenced by the imbalance of the feature's categorization power and redundancy. The paper also offers an examination of the suggested strategy with state-of-the-art methodologies versus benchmark datasets such as the KTH and Weizmann datasets. The study demonstrates that the recommended technique is not just more precise for particular datasets but beats rival methods by an overall weighted average accuracy of 98%. Additionally, the entry offers information on the author's research interests, publications, and accolades.

[16] This study analyzes five independent video activity detection approaches utilizing two publicly available video datasets, the Breakfast collection and the VIRAT 2.0 collection. The authors examine the issues of differentiating normal and abnormal occurrences in long and sophisticated movies with multiple sub-activities and how modifying activity rhythms could influence the performance of

recognition systems. The study offers technical information on the explored video activity detection algorithms and assesses their sensitivity to varied rhythms. The authors also demonstrate that merging related activities in the Breakfast dataset may boost recognition performance for the grouped activity classes and that modifying some of Video Graph's design characteristics may lead to performance increases. Overall, the paper presents a complete overview of video motion detection algorithms with regard to their sensitivity to shifting rhythms when lengthy and complex motion pictures are used.

[17] This paper proposes a deep learning-based methodology for recognizing human activity in videos. The recommended system has three simultaneous tasks that gather information and create classification models based on time, space, and HEVC streaming aspects. The input to the system is a video, which is chronologically separated across 12 equal-sized, distinct parts. The passing in each section is captured using either cumulative picture variations or picture contrasts with motion modification. The apprehend passing is represented as one 2D array, which is combined into an RGB image. The three produced models are merged into a class-grade layer using an average. The output of the algorithm is the finalized label of an individual's activity in the video. The recommended method gives state-of-the-art performance on comparison datasets.

[18] This research describes a system that detects and counts physical activity repetitions from video frames using deep learning methods. The input is a video of a person undertaking a physical activity. The system leverages a 3D posture estimate model to extract the 3D coordinates of 49 human body joints for each frame of the video. The system processes this data using key point normalization and geometry assessment. The system is split into three modules: pre-processing of data, recursive action separation and tally, and event identification. The system utilizes deep semantic features and a recurring segmentation approach to recognize and count the repetitions of physical exercises. The output is the number of repetitions of each exercise done by the participant in the video.

[19] This paper provides two algorithms: the Adaptive Activity Cutting Algorithm (AACA) and the Extreme Learning Machine (ELM). The AACA technique focused on the variation among the active and static sections of the various characteristics collected from CSI signal data. It alters the limit flexibly to get the most effective trade-off between the two. The ELM technique exploits the correlation of several antennas on the WiFi device to extract activity-related Doppler shift correlation values, which are used as input data. It is contrasted with standard classifiers like the Hidden Markov Model (HMM), Sparse Auto Encoder (SAE), Back Propagation Neural Network (BP Neural Network), and Long Short-Term Memory (LSTM). The testing settings specify 10 forms of activities, such as kicking and running, and the exploratory data reveal that ELM has great precision. This study gives numerous novel approaches to identifying activity employing WiFi-based Channel State Information (CSI) data. The implementation of the Doppler shift correlation frequency as a classification feature is a unique approach for acquiring characteristics from the CSI data. The use of a commercial WiFi gadget with several antennas to record CSI data is a unique method of collecting CSI data. Overall, the study proposes a full and unique approach for activity detection leveraging WiFi-based Channel State Information (CSI) data.

[20] This work provides a video-analytic in-class student attention evaluation system that makes use of deep-learning-based visual processing to monitor students' learning activities and deliver feedback to teachers in real-time. The system collects face analysis signals to evaluate student attentiveness and possible tiredness, and stance and activity analysis cues to infer concentration conditions. The solution was verified and assessed using available datasets and a freshly gathered in-class student activity dataset. The authors argue that the approach provides teachers with situational awareness and may help guarantee that students are completely engaged in the learning process.

III. Methodologies And Approches

According to our study, we have analyzed that the algorithms which are mostly used to detect the activities of human are Deep Learning, CNN, LSTM, RF, Logistic Regression, InceptionV3, Xception, InceptionResNetV2. The video datasets are first processed to extract deep features and algorithms are applied on them. Based on the extraction of features from the data sets they classify the activities. In video classification they have also used skeleton joint points for feature extraction and recognition of activities, as it increases the accuracy of the activity recognition. Different datasets have different edge ranges of classification. Activity classification on complex datasets by using these algorithms gives accurate prediction.

IV. Result Analysis

Effective result analysis is a vital component of building and sustaining trustworthy activity detection. It guarantees that the model functions effectively in reality and may assist in improved identification of activities. The following are the usually utilized parameters for the assessment of models:

Accuracy is a common inspection gauge, especially applied to classification tasks. It quantifies the fraction of accurately anticipated cases in the total number of instances. It's a straightforward and obvious measure, but it may not always be the ideal option, especially in circumstances when the classes are unbalanced or the prices of various sorts of mistakes vary dramatically.

$$accuracy = \frac{T.P. + T.N.}{T.P. + T.N. + F.P. + F.N.}$$

Precision in the context of measurement or data analysis denotes the amount of exactness and accuracy in evaluating outcomes. It assesses how closely separate measurements or data points match each other. High precision suggests minimal variability and consistency in data, whereas low precision indicates increased unpredictability and probable mistakes.

$$Precision = \frac{T.P.}{T.P. + F.P.}$$

Recall, also known as True Positive Rate, primarily useful for classification tasks. It assesses the capacity of a model to detect all relevant occurrences within a dataset. It assesses the model's potential to detect all positive.

$$Recall = \frac{T.P.}{T.P. + F.N.}$$

The F1 score is a regularly utilized statistic, particularly when working with unbalanced datasets or when there's a need to balance precision and recall. It's the harmonic mean

of recall and precision and delivers a unique score that takes both false positives and false negatives into consideration.

$$f1 - score = \frac{2 \times precision \times recall}{precision + recall}$$

Here, True Positive or right positive predictions produced, are denoted by the symbol T.P in the formula. False Positive or Inaccurate Positive forecasts are denoted by the letter F.P.

Table – 1: Performance evaluation for Human Activity Recognition on several methods

Technique	precision	recall	F1 score	Accuracy
Softmax	0.5362	0.8627	0.6186	0.4785
OpenMax	0.5362	0.8627	0.6186	0.4785
ODN	0.7032	0.8359	0.7100	0.6185
P-ODN	0.8044	0.7872	0.7690	0.7578
CAC	0.5440	0.8720	0.6289	0.5009
EfficientDetD7	0.80	0.75	0.74	0.75
EfficientDetD4	0.69	0.65	0.62	0.65
DenseNet-121	0.90	0.89	0.89	0.97
MLP	0.93	0.94	0.94	0.97

Table – 2: Accuracy of several models on several datasets

Algorithm	Dataset	Accuracy
Bidirectional-gated recurrent unit (Bi-GRU)	YouTube 11	93.38%
	HMDB5	71.89%
	UCF101	91.79%
Ensemble Model (InceptionV3, Xception, EfficientNetB7, InceptionResNetV2)	Image Datasets	93.76%
Complex Value - Convolutional Neural Networks (CV-CNN)	Motion Capture Radar data	99.81%
Convolutional Neural Networks (CNN)	Florence 3D Action Video Data	93.18%
Convolution Neural Networks-Gated Recurrent Unit (CNN-GRU)	WISDM	96.41%
	UCIHAR	96.67%
	PAMAP2	96.25%
Mahalanobis distance-based mixup triplet learning (MTMD)	UCIHAR	81.22%
	USCHAD	71.25%
	PAMAP2	83.00%
Vision Transformer (ViT) and the Recurrent Transformer (ReT)	UCF101	97.1%
	HMDB52	78.4%
Hierarchical processing and classification pipeline	Imaging Radar Data	87.1%
Convolutional Neural Networks (CNN)	UAVGesture	95%
	Drone Action	90%
	UAVHuman	40%
Logistic Regression	UP Fall	91.51%
	Rzeszow Fall	92.98%
	Sis Fall Data	90.23%
Random Forest	Video dataset	98%
Convolutional Neural Network (CNN)	UCF-ARG aerial dataset	97%
Probabilistic Graphical Models (PGMs), Latent Structured Support Vector Machines (LSSVM)	CAD-60	95.5%
	UT-Kinect	98.3%
	Florence 3-D	87.5%
Multilevel contextual model	VIRAT 2.0	74.41%
(SVM model with standard Radial Basis Function (RBF)	Ground	
	VIRAT 1.0 Ground	60.00%
	UT-Interaction	91.00%
Normalized mutual information-based feature selection (NMIFS)	KTH action dataset	99.0%
	Weizman action	98.22%
Long Term Recurrent Convolutional Networks (LRCN)	Video Dataset	96.27%
Fusion of Deep Learning Algorithms	Video Dataset	96.5%
Deep Learning Algorithms	NOL-18	96.27%
Adaptive Activity Cutting Algorithm (AACA), Extreme Learning Machine (ELM)	WiAct Dataset	94.20%
CNN	In-Class data	80%

V. Future Research Direction

The main challenges addressed by the authors is with the size of the datasets. More number of features will lead to accurate predictions for an activity. In future the author suggested to implement by using other methodologies like RNN, LSTM, YOLO for high accuracy of recognition of human activities. Finally, the proposed models for human activity recognition can be extended to student activity recognition in a classroom by considering various parameters in a video dataset.

VI. Conclusion

In conclusion, this paper explored about different methodologies like Deep learning, CNN, Bi-GRU, RF, Logistic Regression for feature extraction on different data sets which was collected from sensors, wifi-signals, aerial, video data sets. These methodologies, derived from our researched and the literature surveyed, offered educators and researchers valuable tools for assessing student activities and engagement in real-time or through post-analysis. By focusing on the critical task of student activity recognition in the classroom, we had demonstrated the potential of these advanced technologies to provided educators with invaluable insights into classroom dynamics and student engagement. we had strived to create a robust and efficient toolkit for the recognition of student activities. Our approached considers factors such as faced orientation, gazed direction, and head posture to classify students as attentive or non-attentive, offering a nuanced understanding of classroom behavior.

References

[1] Ahmad, Tariq, Jinsong Wu, Hathal Salamah Alwageed, Faheem Khan, Jawad Khan, and Youngmoon Lee. "Human Activity Recognition Based on Deep-Temporal Learning Using Convolution Neural Networks Features and Bidirectional Gated Recurrent Unit With Features Selection." *IEEE Access* 11 (2023): 33148-33159.

[2] Hirooka, Koki, Md Al Mehedi Hasan, Jungpil Shin, and Azmain Yakin Srizon. "Ensembled transfer learning based multichannel attention networks for human activity recognition in still images." *IEEE Access* 10 (2022): 47051-47062.

[3] Yao, Xin, Xiaoran Shi, and Feng Zhou. "Human activities classification based on complex-value convolutional neural network." *IEEE Sensors Journal* 20, no. 13 (2020): 7169-7180.

[4] Endang Sri Rahayu, Eko Mulyanto Yuniarno, I Ketut Eddy Purnama, and Mauridhi Hery Purnomo.Human activity classification using deep learning based on 3D motion feature – ELSEVIER 2023.

[5] Lu, Limeng, Chuanlin Zhang, Kai Cao, Tao Deng, and Qianqian Yang. "A multichannel CNN-GRU model for human activity recognition." *IEEE Access* 10 (2022): 66797-66810.

[6] Lee, Minjung, and Seoung Bum Kim. "Sensor-Based Open-Set Human Activity Recognition Using Representation Learning With Mixup Triplets." *IEEE Access* 10 (2022): 119333-119344.

[7] Wensel, James, Hayat Ullah, and Arslan Munir. "ViT-ReT: Vision and Recurrent Transformer Neural Networks for Human Activity Recognition in Videos." *IEEE Access* (2023).

[8] Zhao, Yubin, Alexander Yarovoy, and Francesco Fioranelli. "Angle-insensitive human motion and posture recognition based on 4D imaging radar and deep learning classifiers." *IEEE Sensors Journal* 22, no. 12 (2022): 12173-12182.

[9] Azmat, Usman, Saud S. Alotaibi, Maha Abdelhaq, Nawal Alsufyani, Mohammad Shorfuzzaman, Ahmad Jalal, and Jeongmin Park. "Aerial Insights: Deep Learning-based Human Action Recognition in Drone Imagery." *IEEE Access* (2023).

[10] Hafeez, Sadaf, Saud S. Alotaibi, Abdulwahab Alazeb, Naif Al Mudawi, and Wooseong Kim. "Multi-sensor-based Action Monitoring and Recognition via Hybrid Descriptors and Logistic Regression." *IEEE Access* (2023).

[11] Ramirez, Heilym, Sergio A. Velastin, Ignacio Meza, Ernesto Fabregas, Dimitrios Makris, and Gonzalo Farias. "Fall detection and activity recognition using human skeleton features." *Ieee Access* 9 (2021): 33532-33542.

[12] Aldaihoul, Nouar, Hezerul Abdul Karim, Aznul Qalid Md Sabri, Myles Joshua Toledo Tan, Mhd Adel Momo, and Jamie Ledesma Fermin. "A comparison between various human detectors and CNN-based feature extractors for human activity recognition via aerial captured video sequences." *IEEE Access* 10 (2022): 63532-63553.

[13] Arzani, Mohammad M., Mahmood Fathy, Ahmad A. Azirani, and Ehsan Adeli. "Switching structured prediction for simple and complex human activity recognition." *IEEE transactions on cybernetics* 51, no. 12 (2020): 5859-5870.

[14] Ajmal, Muhammad, Farooq Ahmad, Mudasser Naseer, and Mona Jamjoom. "Recognizing human activities from video using weakly supervised contextual features." *IEEE Access* 7 (2019): 98420-98435.

[15] Siddiqi, Muhammad Hameed, Madallah Alruwaili, and Amjad Ali. "A novel feature selection method for video-based human activity recognition systems." *IEEE Access* 7 (2019): 119593-119602.

[16] Ayhan, Bulent, Chiman Kwan, Bence Budavari, Jude Larkin, David Gribben, and Baoxin Li. "Video activity recognition with varying rhythms." *IEEE Access* 8 (2020): 191997-192008.

[17] Shanableh, Tamer. "ViCo-MoCo-DL: Video Coding and Motion Compensation Solutions for Human Activity Recognition Using Deep Learning." *IEEE Access* (2023).

[18] Cheng, Sheng-Hsien, Muhammad Atif Sarwar, Yousef-Awwad Daraghmi, Tsì-Uí İk, and Yih-Lang Li. "Periodic physical activity information segmentation, counting and recognition from video." *IEEE Access* 11 (2023): 23019-23031.

[19] Yan, Huan, Yong Zhang, Yujie Wang, and Kangle Xu. "WiAct: A passive WiFi-based human activity recognition system." *IEEE Sensors Journal* 20, no. 1 (2019): 296-305.

[20] Su, Mu-Chun, Chun-Ting Cheng, Ming-Ching Chang, and Yi-Zeng Hsieh. "A video analytic in-class student concentration monitoring system." *IEEE Transactions on Consumer Electronics* 67, no. 4 (2021): 294-304.