[1]Ananda Ravuri

[2]Dr.Siva Shankar S

[3]D. P. Devan

[4]Manoj Kumar Padhi

[5]Dr. V K Senthil Ragavan

[6]Dr. Mahesh Maurya

[7]Dr.A.Ravi

A Systematic Literature Review on Human Activity Recognition

Abstract: - Human Activity Recognition (HAR) plays a significant role in several fields by automatically identifying and monitoring human activities using advanced techniques. It enhances safety, improves healthcare services, optimizes fitness routines, and enables context-aware applications in various fields. HAR contributes to a more efficient and intelligent interaction between humans and technology. It has emerged as an essential research domain with applications in healthcare, smart environments, and human-computer interaction. This study aims to provide a comprehensive survey of the evolving landscape of HAR, including key methodologies, techniques, and trends in existing research. The study discusses various applications of HAR and their significance in modern smart environments. The survey also highlights different types of HAR and data collection techniques. Additionally, it explores various methods for analyzing the collected data and provides a comprehensive analysis of existing human activity classification datasets. It offers valuable insights into understanding the strengths and limitations of various HAR techniques. The study also discusses various challenges and future directions for HAR.

Keywords: Human Activity Recognition, Activity Monitoring, Machine learning, Vision, Sensors, Human-Centric Sensing, Deep Learning, Healthcare, Performance Recognition

[1] [1]Senior Software Engineer, Intel corporation, Hillsboro, Oregon 97124 USA.

Email :Ananda.ravuri@intel.com

[2]Associate Professor & Head IPR, Department of CSE, KG Reddy College of Engineering and Technology, Hyderabad, Telangana, India – 501504.

Email :drsivashankars@gmail.com

[3]Assistant Professor, Department of Computer Science & Engineering, K. Ramakrishnan College of Technology (Autonomous), Kariyamanickam Road, Samayapuram, Trichy.

Email :devan.p.durairaj@gmail.com

[4]Assistant Professor, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Green Fileds, Vaddeswaram, A.P. – 522302.

Email :manojpadhi1503@gmail.com

[5]Professor and Head, Department of Information Technology, St. Martin's Engineering College,

Dhulapally, Kompally, Secunderabad-500100, Telangana, India.

Email :vksenrag@yahoo.com

[6]Associate Professor and HoD, K. C. College of Engineering, Department of Computer Engineering, Mith Bunder Road, Near Hume Pipe, Kopri, Thane (E) - 400603.

Email :mahesh.maurya@kccemsr.edu.in

[7]Professor of ECE & HOD, PSCMR College of Engineering and Technology, Vijayawada,AP,India.

Email :ravigate117@gmail.com

Copyright © JES 2024 on-line : journal.esrgroups.org

1. Introduction

The process of identifying and comprehending diverse tasks conducted by persons is referred to as human activity recognition [1]. This recognition is critical in human-to-human interaction and interpersonal relationships due to its complexity while conveying important

information about a person's identity, personality, and psychological condition [2]. Humans' ability to identify and interpret the actions of others is a major subject of research in domains such as computer vision and machine learning. Several applications have been developed as a result of substantial research in these areas. Video surveillance systems, for example, benefit from identifying human actions to improve security and situational awareness [3]. Understanding human actions also helps in the creation of more intuitive and responsive interfaces in the field of human-computer interaction. Recognizing human actions is critical in robotics for defining human behavior and allowing robots to interact with humans in a more natural and adaptive manner. In conclusion, human activity recognition is critical in many sectors, enabling the development of sophisticated systems and technologies that improve human-computer interactions, increase security, and enable robots to properly perceive and respond to human activities.

Human Activity Recognition (HAR) is the process of identifying and comprehending human actions from diverse sources such as photographs or sensor data. Despite its importance, HAR is a difficult process due to a variety of difficulties such as background clutter, partial occlusion, size and viewpoint shifts, illumination fluctuations, and intra-class variances. Furthermore, annotating behavioral roles takes time and requires event knowledge. Deep learning, feature engineering, sensor fusion, transfer learning, and temporal modeling are examples of HAR approaches. HAR has applications in a wide range of sectors, including healthcare, smart homes, sports, security, and human-computer interaction. Despite the difficulties, continuous research in this area is improving the accuracy and usability of HAR systems in real-world circumstances.

1.1 Human Activities Levels

Human activities exhibit a diverse range, varying from simple gestures to complex group interactions, as illustrated in Figure [4]. To effectively train machines for Human Activity Recognition (HAR), these activities are categorized into five distinct types or levels based on their complexity and duration. The ultimate goal of HAR is for systems to be able to accurately recognize and differentiate between those actions. Each activity category is described in detail below, including human-object interaction as well as human-human relationships merged under the term interaction.

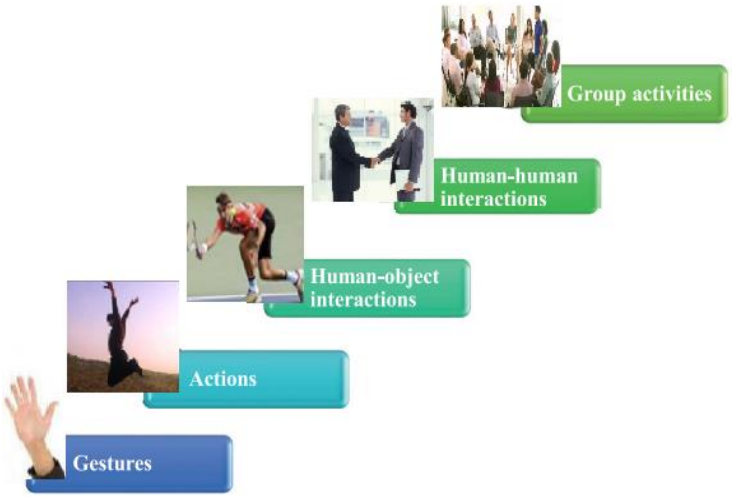

Figure 1: Different Levels of human activities [4]

- Gesture: Gestures involve basic hand movements or motions made by using other parts of the individual's body to communicate specific ideas or meanings. Examples include head shakes facial expressions, and hand-waving. Gestures are typically brief and represent the simplest form of activity among the categorized types.
- Action: Actions are relatively straightforward activities performed by humans, involving a sequence of gestures. Examples encompass actions such as knocking, swimming, and running. Actions are more complex than gestures and often comprise multiple coordinated movements.
- Interaction: Interactions entail activities carried out by two agents, where one agent is human, and the second agent is either a person or an object. Depending on the nature of the agents involved, interactions are further classified into human-object interactions and human-human interactions. Human-human interactions include activities like hugging, shaking hands and wrestling, whereas human-object interactions involve interactions between a person and entities like mobile phones or laptops.
- Group Activity: Group activities represent the most intricate form of human activities, involving the participation of more than two individuals and potential interaction with one or multiple objects. These activities consist of a series of actions, gestures, in addition to interactions. Examples include group study sessions, football matches, and presentations. Group activities encompass complex social dynamics and require recognition of various gestures, actions, and interactions within a collective setting.

Understanding these levels of human activities forms the foundation for developing sophisticated HAR systems. By discerning the nuances between different activity types, researchers and developers can create intelligent algorithms and models capable of accurately identifying and categorizing human activities in diverse contexts.

1.2 AI in HAR

Artificial intelligence (AI) has undergone rapid advancement, revolutionizing numerous domains such as intelligent video surveillance systems, self-driving vehicles, assisted living, and human-computer interface systems. A fundamental task central to these applications is video-based human action recognition, a domain that has witnessed continuous evolution and improvement. The primary objective of human action recognition is to decipher human activities from video data [5]. Current research in this field explores diverse avenues, including the effective fusion of multi-modal information [6], learning without annotated labels [7], training models with limited data points [8] and the exploration of novel architectural designs [9].

1.3 Applications of Human Action Recognition

HAR is used in a variety of applications, including intelligent video surveillance systems, self-driving vehicles, assisted living, human-computer interface systems, healthcare and rehabilitation, sports and fitness, and gaming and entertainment. AI-powered HAR improves video surveillance capabilities by providing real-time analysis of human actions in security footage [10]. This technology aids in the detection of suspicious activity, the protection of the public, and the prevention of security breaches. HAR is critical in self-driving vehicles because it allows them to perceive and respond to human actions and gestures [11]. This skill guarantees that autonomous vehicles interact safely with pedestrians, bicycles, and other drivers on the road.

HAR is used to monitor the actions of elderly or disabled people in assisted care facilities [12] [13]. AI-powered devices can detect falls, track movements, and provide prompt assistance, improving residents' quality of life while maintaining their safety. Human-computer interactions are made more natural and intuitive by HAR. AI-powered interfaces provide hands-free control of devices by recognizing gestures, facial expressions, and body motions, resulting in a more smooth and user-friendly experience. HAR is used in healthcare to monitor patients' physical activities and rehabilitation programs [14]. AI algorithms assess patients' actions to ensure they follow recommended regimens, assisting healthcare providers in analyzing progress and changing treatments.

HAR is used to measure athletes' movements [15] in sports training and fitness applications, delivering significant insights for performance enhancement. AI systems examine approaches, allowing instructors and players to fine-tune their talents and avoid injury. HAR improves user experiences in the gaming industry by providing gesture-based controllers and immersive interactions. Natural movements can be used by players to interact with games, producing a more dynamic and engaging gaming environment. Thus, AI-powered HAR not only improves the efficiency and safety of numerous industries, but also promotes creativity, providing the path for unique

applications and revolutionary user experiences across multiple disciplines.

2. Literature Survey

The evolution of Human Activity Recognition (HAR) technology has been instrumental in processing data sourced from an array of sensing devices, ranging from vision sensors to embedded sensors[16]. This advancement has spurred the creation of context-aware uses, particularly within evolving fields like the Internet of Things (IoT) as well as healthcare. Although previous studies have explored specific aspects of HAR, this comprehensive review takes a holistic approach. It delves into key themes within HAR and scrutinizes the most recent advancements in the field[17]. The review categorizes HAR methodologies into two primary groups including sensor-based as well as vision-based HAR, contingent on the type of data generated. Subsequently, it conducts an in-depth analysis encompassing datasets, pre-processing techniques, feature engineering, as well as the training process, particularly focusing on the integration of deep learning methodologies in HAR. The study not only highlights the current challenges faced in HAR but also provides valuable insights and recommendations for future research directions, ensuring a nuanced understanding of this dynamic field[18].

2.1 Advances in Sensor Technology and Human Activity Recognition

In the previous years, sensor technology has made remarkable strides in several areas including computational power and manufacturing costs with accuracy and size[19]. These developments have facilitated the integration of various sensors into smartphones and portable devices, enhancing their intelligence and utility. Simultaneously, the evolution of video surveillance technology resulted in improvements in the quality of video, simplified arrangement, reduced costs, and secure communication [20]. Consequently, a cumulative amount of applications applying Closed-Circuit Television (CCTV) systems in monitoring and security systems have emerged [21].

While sensors serve specific purposes, they universally collect raw data from their surroundings, which, when analyzed, yields valuable insights. HAR enables machines to monitor and grasp diverse human actions using sensors along with multimodal input data. [3]. Early HAR efforts, dating back to the 1990s, achieved over 95% accuracy in controlled data collection scenarios [7]. With the rapid advancements in smartphones, wearables, and CCTV systems, researchers have been motivated to enhance HAR systems for practical applications.

HAR finds applications in behavior analysis [10], surveillance systems [8,9], gesture recognition [11–13], Ambient Assisted Living (AAL) [16,17], various healthcare systems [18,19] and patient monitoring [14,15]. For instance, patients with chronic conditions need to adhere to specific diets and exercise regimens [20]. Continuous activity tracking provides real-time feedback to patients and enables clinicians to monitor progress. Similarly, patients with cognitive impairments require continuous monitoring to detect unusual actions promptly, preventing adverse outcomes [21]. In tactical situations, real-time response to the actions of soldiers, their positions, as well as essential signs is crucial for skill development and safety. This feedback also aids commanders in decision-making during training and combat scenarios [22].

However, HAR poses challenges due to the lack of standardized procedures for associating collected data with specific actions. Additionally, managing the substantial volume of collected data remains a significant technical challenge. Researchers continue to explore innovative solutions to address these complexities, striving to enhance the accuracy, efficiency, and applicability of HAR systems in various real-world contexts.

2.2 Wearable Sensor HAR

Wearable sensor technology entails integrating intelligent electronic devices into wearable items or directly onto the body, allowing for the measurement of various biological and physiological signals such as blood pressure, heart rate, accelerometers, body temperature, and attributes such as location and motion. The sensors exchange data with assimilation devices such as cellphones, laptop computers, or specialized embedded systems. The accumulated raw signals are routed to application servers enabling continuous monitoring, visualizing, and evaluation [17]. While smartphones with gyroscopes, cameras, and accelerometers can help with activity recognition, their efficiency is restricted for complicated activities [9]. As a result, more sensors or specialized sensing devices are required to improve recognition accuracy. Numerous machine learning techniques have been presented in the literature for processing features acquired from raw data in order to detect human actions. Existing

approaches, on the other hand, frequently rely largely on heuristic handcrafted feature extraction [22], restricting their application. This review looks at new approaches that use wearable sensor technologies and powerful machine learning algorithms to recover the accuracy and variety of human activity recognition systems.

Figure 2: Wearable sensors and devices

Several studies have explored sensor-based HAR, considering it as a time series classification task utilizing data from diverse sensor devices like electrocardiography (ECG), inertial measurement units (IMU), heart rate measurement (HRM) and electromyography (EMG), [37]. A general-purpose framework by Bulling et al. treats individual sensor data frames as statistically independent [38]. Traditional HAR approaches in addition to deep learning-based techniques categorize existing research. To maintain spatial data in signal frames, Hammerla et al. [41] presented an empirical cumulative density function (ECDF) attribute. This method is protracted by Kwon et al. [9] incorporating temporal structures into ECDF, enhancing activity recognition. Anguita et al. [40] utilized smartphone-based model with multi-class SVM for six-class locomotion activity recognition. Bhattacharya and Lane [45] also presented a model optimization strategy for wearable devices constrained by resources. Recurrent deep learning methods, especially LSTM units, have excelled in HAR, with Ordóñez and Roggen [11] introducing DeepConvLSTM, combining LSTM with CNN layers for capturing short-term as well as long-term sequential associations. Deep learning methods [42, 43, 44] have gained traction, with Yang et al. [10] proposing a CNN-based approach for human activity recognition, employing manifold convolutions as well as pooling filters beside sequential magnitudes [10]. Alsheikh et al. [44] developed a hybrid technique employing deep belief networks in conjunction with hidden Markov models for feature extraction.

2.3 Subject-independent HAR

In addressing interpersonal variability in HAR, researchers have explored various strategies. One approach involves increasing the training data from diverse subjects, although this is often impractical due to the cost and complexity of collecting and annotating data from diverse individuals. To mitigate this challenge and overcome cross-domain HAR difficulties, transfer learning algorithms were explored encompassing cross-subjects [54][55], cross-locations[53] and cross-sensor-modalities [52]. Domain adaptation, a subdivision of transfer learning, measures and aligns heterogeneity of data circulation [56]. Transfer learning utilized by Zhao et al. [58] includes embedded decision tree algorithm, integrating decision trees with k-means clustering for recognition of personalized activities based on mobile phone data. A cross-person activity recognition by Deng et al. utilizes a condensed kernel extreme machine learning, classifying target samples and high-confidence samples, and then integrating them into the training dataset [57].

Recent advancements include multi-tasking in addition to GAN approaches (generative adversarial learning) to tackle issues in the distribution of diverse data. Chen et al. [18] presented the METIER model, a deep learning-based multi-task approach for user recognition and activity. This model shares parameters between user recognition and activity modules, improving activity recognition through a common attention device. Sheng et al. [19] introduced inadequately supervised multi-task depiction, utilizing Siamese networks as well as a temporal backbone model using the convolutional network. According to Bai et al. [20], a discriminative adversarial multi-view network employing CNN extract multi-view feature utilizing WGAN and Siamese network design to reduce deviations between various subjects' features.

Additionally, attention and transformer-based models [50], well-known for their success in skeleton-based HAR [32], natural language processing [67], and computer vision [31], have been applied to enhance HAR techniques [31][32]. Particularly, Miao et al. [69] introduced DynamicWHAR, a framework using GCN for dynamic inter-sensor correlation learning. Dynamic features were extracted from multi-sensor information through modeling the dynamic connections between distinct sensors.

2.4 Human Activity Recognition Enhancing Techniques

1. Machine learning and deep learning

Signal processing technology has traditionally been critical in interpreting raw data from sensors [66], whereas computer vision (CV) technology has been used for preprocessing and extracting handcrafted features from images or movies [9]. These techniques were critical in feature engineering because they generated sensor-specific or signal-specific features, which were then trained using machine learning (ML) algorithms to make classification decisions. However, the human examination of datasets to identify acceptable features, as well as the subsequent feature extraction procedure, caused difficulties, particularly when dealing with new datasets or sensors, resulting in complexity and lack of scalability [67].

Deep learning has evolved as a dominating study area in recent decades, reaching human-level performance in a variety of domains, including HAR. Deep learning excels at handling large datasets by spontaneously extracting abstract features from sensor inputs or image sequences, removing the need for manual feature engineering. This breakthrough beat previous machine learning methods that relied on handcrafted and domain-related features [12][68]. Consequently, deep learning has paved the way for novel HAR solutions, allowing for the use of bigger datasets and the construction of real-time systems of HAR.

Hassan et al. introduced a HAR framework using deep learning based on smartphone sensors, outperforming traditional multiclass ML techniques such as Support Vector Machines (SVM) as well as artificial neural networks (ANN) [69]. A study [70] used a convolutional neural network model (CNN) for extracting simple statistical and local data, attaining advanced real-time performance at a reduced computational rate. Furthermore, a method using deep bimodal learning for detection of human fatigue expression demonstrated excellent results, beating earlier algorithms including a recognition rate of more than 96% [71].

Deep learning advances have had a tremendous impact on the field of HAR, resulting in improved performance, and the development of real-time systems using greater datasets. In terms of learning algorithms, feature engineering, and data pre-processing, the important alterations between deep and learning machine learning approaches in HAR highlight distinct advantages and challenges in each methodology. In traditional machine learning, data pre-processing is crucial and demands meticulous normalization techniques for optimal performance. Feature engineering relies heavily on manual extraction, often requiring tailored approaches for specific applications. This method struggles with the intricacies of complex activities and frequently needs additional processes like dimensionality reduction and feature selection. On the contrary, deep learning eliminates the need for intricate data pre-processing, as it can automatically learn abstract features from input including raw data. Deep learning techniques excel in identifying temporal and spatial, dependencies with measure invariant features autonomously, making them well-suited for intricate and varied datasets. While machine learning can operate effectively with smaller training datasets and consumes limited computational resources, deep learning requires extensive datasets to avoid overfitting and involves high computational complexity. Furthermore, particular hardware is often necessary to hasten the training procedure in deep learning applications. Thus, the

trade-offs between the two approaches emphasize the need for careful consideration based on the specific requirements and complexities of the HAR.

2. Ensemble Learning

The rapid advancement of deep learning techniques has significantly influenced HAR, offering efficient feature learning capabilities through its layered structure. However, deep learning models often require substantial labeled data, which can be challenging to obtain. Moreover, their high computational demands hinder real-time activity detection [8]. To address these challenges, some studies have explored the integration of ensemble learning algorithms with traditional machine learning classifiers like Multilayer Perceptron, Logistic Regression, Support Vector Machine, K-Nearest Neighbor, and Random Forest. They employed a voting algorithm combining the strengths of all classifiers in HAR. The technique was introduced in a previous study [7], utilizing more efficient base classifiers and demonstrating superior performance and effectiveness in improving HAR accuracy while simplifying implementation.

The complex deep models come with significant demands on computational power and memory resources. Additionally, existing studies make a common assumption that the data distributions among subjects in training as well as testing datasets are identical. In practical HAR scenarios, where individuals tend to perform the same activities in diverse manners, there are substantial disparities in data distributions among subjects.

2.5 Hybrid sensors

In recent years, researchers have endorsed the use of hybrid sensors, which combine different sensor types, to improve the accuracy and durability of HAR systems [18,69]. This novel approach has helped to improve the recognition rates of complex indoor activities. As illustrated in the image, a hybrid sensors framework was designed in [97] to successfully recognize complicated 21 activities in an indoor scenario. This approach comprises three separate sensing contexts such as body sensing, location sensing, and environmental sensing.

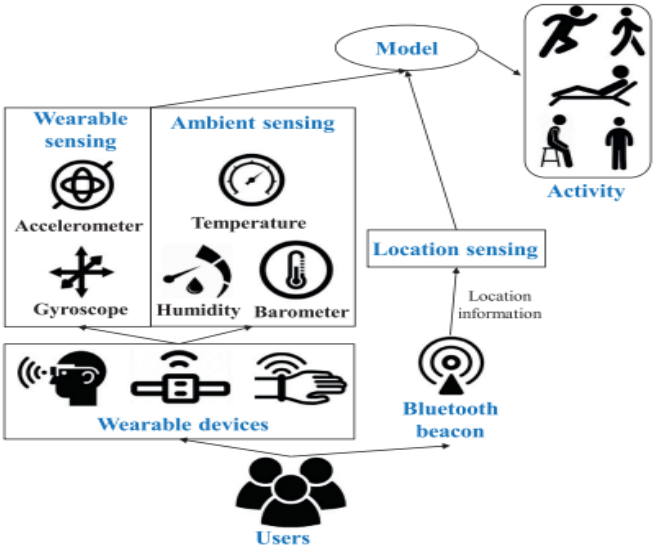

Figure 3: HAR with hybrid sensor

Wearable sensors that gather data about human movements and physiological signals are at the heart of body sensing. Environmental sensing combines data from sensors placed in the environment to provide insights into the conditions. Object sensors are used in location sensing to discern spatial information. Researchers are working on creating a complete and robust HAR system capable of reliably recognizing a wide range of actions in varied circumstances by merging these distinct sensor types. This method emphasizes the importance of hybrid sensors in furthering the field of HAR and solving the obstacles provided by complex activity identification scenarios [18,69].

The obtained data is fed into a model for human activity classification, which includes position data gathered from

object sensors, while a mix of wearable and environmental sensors is used to observe human actions and perceive the surrounding environment.

The table below summarizes the key characteristics of frequently used sensor-based HAR datasets. Wearable sensors like accelerometers, gyroscopes, and magnetometers are often used for training models. WISDM and UCI datasets are standard datasets that are frequently employed for assessing sensor-based HAR models. They were recently applied to assess the performance of deep learning models. The WISDM database approached 93% prediction accuracy, whereas the UCI dataset's classification accuracy topped 97% [70].

Table 1: publicly available sensor-based HAR datasets

Dataset	**Activities**	**Attr.**	**Devices**	**Sampling rate Hz**	**Sensors**
HASC [101]	6	4	Smartphone	10-100	Accelerometer Gyroscope Magnetometer GPS
UCI HAR [104]	6	561	Smartphone	50	Accelerometer Gyroscope
UniMiB SHAR [98]	17	NA	Smartphone	1-32 K	Accelerometer
UCI Heterogeneity AR [100]	5	16	Smartphone & wearable sensors	100-200	Accelerometer Gyroscope
Real world [99]	8	7	Smartphone & wearable sensors	50	Accelerometer
WISDM [106]	6	46	Wearable sensors	20	Accelerometer
UCI M-HEALTH [102]	12	23	Wearable sensors	50	Accelerometer Gyroscope Magnetometer, GPS
UCI OPPORTUNITY [105]	6	242	Wearable sensors	NA	Hybrid
UCI AR-HOP [103]	7	9	RFID	NA	Object sensor

3. Vision-based HAR

Over recent decades, Human Activity Recognition using vision-based systems has extended substantial attention owing to its applications in real-world scenarios such as Closed-Circuit Television (CCTV) systems deployed in public areas, enhancing surveillance and safety measures. Studies on vision-based systems can be categorized according to the type of data utilized, which comprises RGB [12,107,108] as well as RGB-D data [9,25]. Studies have shown that these frameworks employing RGB-D data tend to achieve higher accuracy compared to those relying solely on RGB data [12,109]. This advantage arises from the additional information and depth channels provided by multi-modal data. However, despite the improved accuracy, RGB data are widely preferred in existing frameworks of HAR due to the challenges associated with computational loads for large datasets, higher prices and configuration intricacy, involved in implementing RGB-D data solutions in practical settings.

The challenges posed by issues such as partial occlusion, background clutter, variations in scale, lighting, viewpoint, as well as appearance make it difficult to recognize human activities from video sequences or still images. A multiple motion recognition technique is required for numerous uses, involving human-computer interfaces, video surveillance systems, as well as robotics in behavioral character analysis. Human activity recognition techniques can be broadly classified into binary groups based on the use of data from various modalities. Furthermore, it is crucial to conduct a thorough examination of existing publicly available human activity classification datasets.

Human Activity Recognition (HAR) datasets are critical for developing intelligent systems that can understand

and respond to human behaviors. These datasets provide critical information about three main aspects: the sensor device, the subject or actor, and the sensing environment. The dynamic and malleable character of these parameters, however, classic machine learning premise confronts the same domain of source and target data. This problem has been efficiently handled by the knowledge transfer paradigm, which has transformed the field. Knowledge transfer removes the limits imposed by this traditional machine learning theory, allowing significant insights learned in one area to be transferred and used in another. Considering the dynamic nature of sensors and settings, one of the key issues in HAR is the appropriateness of earlier training data for real-time recognition. In this setting, transfer learning appears as a powerful option, allowing for the exploitation of existing data samples and their integration into classification, regression, and recognition tasks. This method not only compensates for old data, but it also eliminates the requirement for the complexity of collecting new training data, alleviates a load of costly data labeling efforts, and considerably improves testing data correctness. Several studies investigate the transformational impact of knowledge transfer and transfer learning approaches in the field of Human Activity Recognition, delving into their applications, methodologies, and contributions to real-world settings.

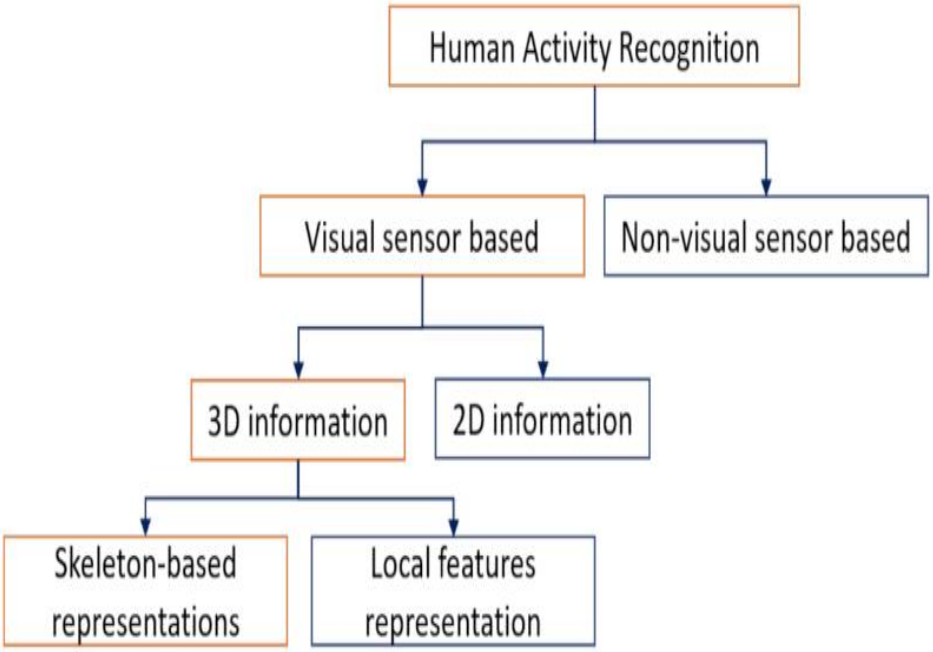

Figure 4. HAR approaches based on the information source

3.1 RGB data

A standard RGB image captures visible light using red, green, and blue channels, detected by cameras using standard CMOS sensors (complementary metal-oxide-semiconductor). For instance, in a study conducted by Zerrouki et al. [108], an effective AdaBoost algorithm was employed for Human Activity Recognition utilizing CCTV video footage. The researchers tested their framework using publicly accessible URFDD and the Universidad de Malaga datasets for fall detection. The outcomes of the experiments demonstrated the model's ability to achieve high classification accuracy when applied to RGB datasets. RGB data are readily accessible, cost-effective, and provide detailed texture information about the subjects. However, it is essential to note that these sensors have limited range, are sensitive to calibration, and are significantly influenced by environmental settings such as illumination, lighting, and the presence of messy backgrounds.

3.2 RGB-D data

With advancements in depth sensors as well as range imaging methods [110], Human Activity Recognition has significantly improved in accuracy. The figure illustrates that, in addition to capturing unique RGB data, depth information is also recorded in RGB-D cameras, enabling algorithms for more precise recognition of human activities. Moreover, from this depth data, skeleton data can be extracted to create a condensed representation of

the skeleton of the human body, as demonstrated in the figure. Skeleton data exist within a low-dimensional space [8], enabling HAR models to operate more swiftly. Utilizing 3D human joint data from depth cameras has become a promising research avenue due to its applicability in various fields.

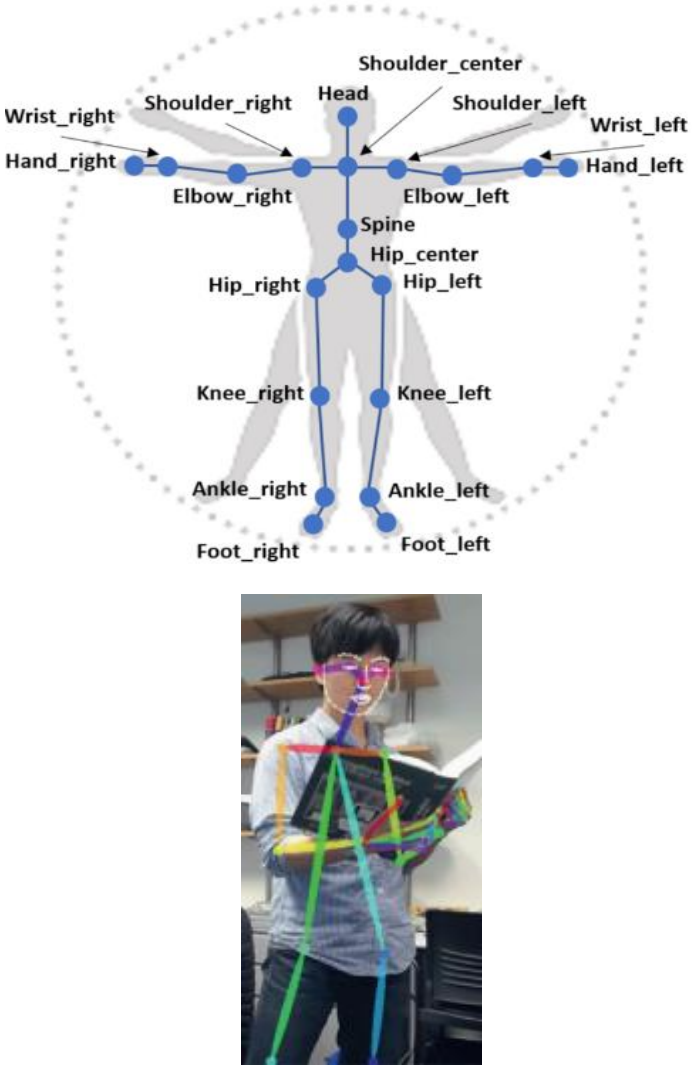

Figure 5: Skeleton model by Microsoft Kinect device and skeleton detection

Cippitelli et al. [112] devised a robust Human Activity Recognition (HAR) algorithm utilizing RGB-D skeleton data. Their model demonstrated exceptional performance on two standard datasets, namely CAD-60 and KARD. In a separate study, Jalal et al. [9] proposed a HAR system using RGB-D sequences captured by a Kinect device, integrating depth silhouettes and human skeletons as key features. Through extensive evaluations of three benchmark depth datasets, their method achieved top-tier results. RGB-D data offers several advantages over RGB data, including resilience to lighting variations, adaptability to illumination shifts, and effectiveness in low-light conditions. Despite these advantages, RGB-D data have limitations such as low resolution, susceptibility to noise due to low sensitivity, and vulnerability to specific materials like light-absorbing and transparent substances.

Table 2: Vision-based HAR - Publicly available datasets

Name	ID	Videos	Activities	Year	Depth
Kinetics-700 [113]	V4	650,000	700	2019	Y
NTU RGB+D [114]	V10	56,880	60	2016	Y
Berkeley MHAD [124]	V13	660	11	2014	Y
UTD-MHAD [122]	V11	861	27	2015	Y
HDM05 [130]	V20	1500	70	2007	Y

SBU Kinect interaction [127]	V16	300	7	2012	Y
CAD-120 [125]	V14	120	4	2013	Y
HACS [115]	V1	1,550,000	200	2019	N
Hollywood2 [129]	V19	3669	12	2009	N
Moments in Time [116]	V2	1,000,000	339	2019	N
UT-Interaction [128]	V18	180	10	2010	N
AVA [5]	V3	430	80	2018	N
Sports-1M[123]	V12	1,100,000	487	2014	N
MultiTHUMOS [117]	V5	400	65	2018	N
UCF101 [126]	V15	13,320	101	2012	N
20BN-something [118]	V6	220,847	174	2017	N
ActivityNet200 [121]	V9	19,994	200	2016	N
Charades-Ego [119]	V7	7860	157	2016	N
HMDB51 [109]	V17	7000	51	2011	N
DALY [120]	V8	8133	10	2016	N

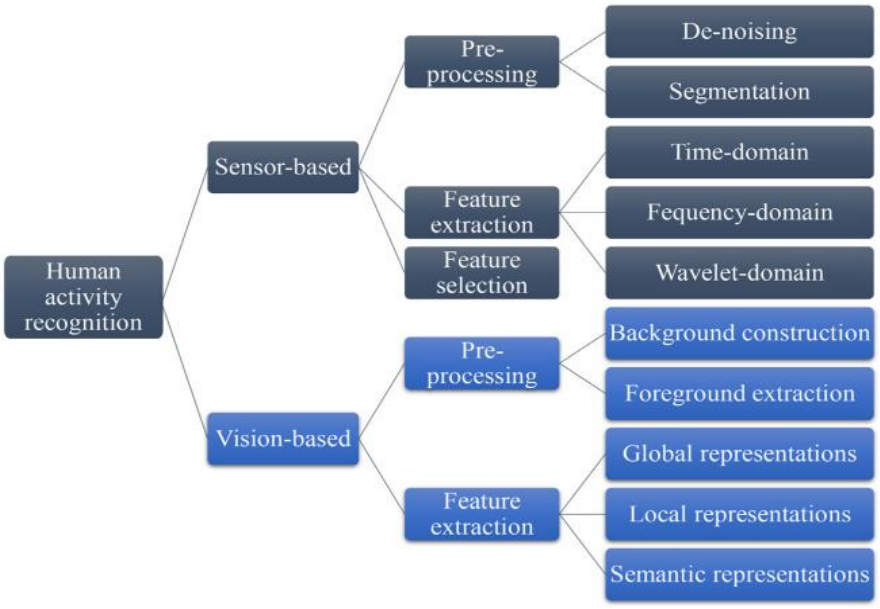

Figure 6: HAR data processing in sensor and vision approach

4. Data processing

Data pre-processing as well as feature engineering are critical steps in preparing raw sensor data for machine learning models in sensor-based Human Activity Recognition (HAR). Denoising techniques are used to reduce noise caused by a variety of causes such as miscalibration and environmental conditions. Noise reduction methods include low-pass filters, mean filters, and wavelet filters [133-135]. Furthermore, segmentation is critical because activities frequently exceed sensor sampling rates. Data streams are divided into segments that map to certain activities using segmentation techniques such as time-driven, event-driven, and action-driven windows [138]. Hammerla et al. presented real-time applications with a 1-second sliding window illustrating a 50% overlay, displaying superior performance in wearable sensor datasets [139]. Then, feature extraction methods are used to extract important information from pre-processed data. According to signal characteristics, time-domain features (TDFs) such as median, variance, and kurtosis, as well as frequency-domain features (FDFs) like spectral entropy and peak power, are extracted [9,141]. Wavelet transform (WT) has gained popularity due to its capacity to calculate wavelet energy in components, which provides useful characteristics for classification [142]. Wrapper, filter, and embedding approaches are used to pick a subset of important characteristics, improving classification

accuracy and decreasing dimensionality [147-149]. These techniques work together to improve sensor data for accurate HAR in a variety of applications.

Diverse activities, lighting conditions, occlusion, complex backgrounds, and viewpoint fluctuations, provide problems in vision-based Human Activity Recognition (HAR), making segmentation and feature engineering critical for enhancing HAR performance [2]. An important procedure in vision-based HAR is segmentation, which includes extracting target subjects from picture or video sequences. Background creation and foreground extraction approaches are separated [150]. Due to shifting backgrounds, foreground extraction approaches, particularly with moving cameras or devices, are preferable. To extract objects, temporal, geographical, or spatiotemporal information is evaluated, giving important data for future feature-based analysis. Methods for extracting features are critical for converting raw visual input into usable information. Handcrafted characteristics, spanning global, local, and semantic techniques, are included in traditional methods [29,137]. Global features extract descriptors from videos or photos. Histogram of Oriented Gradients (HOG) and Discrete Fourier Transform (DFT) techniques were employed. While they are effective for preliminary studies, they struggle with occlusions and changing perspectives. Local features use techniques such as HOG, Scale-Invariant Feature Transform (SIFT), and Speeded-Up Robust Features (SURF) to focus on localized patches. These techniques are resistant to partial occlusions and noise. Semantic characteristics are designed to emulate human perception by including factors such as body postures, scenarios, visual features, and linked items. Pose estimation, appearance-based methods (local and global aspects), and 3D approaches (using depth map data) all help to improve comprehension of human activities. These segmentation and feature engineering procedures are critical in the advancement of vision-based HAR, allowing for accurate recognition in a variety of demanding settings.

4.1 Comparison

Table 3: Comparing various HAR studies and their performances

Sr. No.	Application	Technique	Dataset	Devices	Performance	Details
Vision Based Model						
1	Indoor HAR [199]	5-CNNs	V10, V14, V16	RGB-D	95.11 % (V10) 96.67 % (V16)	Integrates distinct CNN classifiers Novel technique for skeleton data processing
2	Intelligent vehicles [219]	Customized CNN	Self	K	Detection rate of 91%	Customized CNN Recognize behavior of driver Unsupervised segmentation using Gaussian mixture technique
3	HAR[200]	RNN tree	Self, V9	NA	V9-0.832%	Tree structure of multiple RNN models Transfer learning Adaptability for addition of new class Big scale dataset
4	HAR[211]	AE	V20	NA	Better for data corrupted by noise	Denoising tensor autoencoder (DTAE) Temporal corruption – different corruption ratios. Spatial corruption - diverse spatial noise, similar ratio for temporal corruption
5	Hybrid Video streaming[191]	CNN and AE	Self	A, G and M	97.8%	Efficient and optimized Rapid dynamic frame skipping Video data State-of-the-art real-time data from CCTV

6	Hybrid Surveillance [168]	CNN and LSTM	V15	NA	94.4%	CNN-based optical flow Pre-trained MobileNet model
Sensor Based Model						
1	Mobile HAR [70]	Customized CNN	S7, S10	A	S7 -97.62%, S10 -93.32%	Narrow architecture of CNN Statistical and global features Time-series with Real-time data Cutting-edge performance UCI and WISDM HAR datasets
2	RNN Gesture recognition [220]	LSTM	S8, S9	S A, G and M	S9-80%	RNN- hand activities of six categories Inertial sensors' data No pre-processing
3	RNN HAR [221]	Residual Bidir-LSTM	S7, S8	A, G and M	S7 -93.6%, S8 -90%	Increased speed Modifying LSTM Importance of window size for HAR
4	HAR [190]	Continuous AE	Swiss-roll	A, G and M	98.4%	Continuous auto encoder Less training period. Feature extraction for Frequency and Time
5	Fall detection [192]	Customized AE	DLR, COV	A, G and M	Decent balance among FPR and TPR	Different auto-encoder models Wearable devices Threshold tightening method
6	HAR [68]	DBN	S5	A, G and M	97.5%	DBN framework Higher performance
7	HAR [69]	DBN	S5	A, G and M	95.85%	DBN Smartphone inertial sensors Beats traditional approaches Better than ANN and SVM
8	Hybrid Indoor HAR[222]	CNN+LSTM	S8	I and 3-axis A	91%	Deep learning Convolutional and LSTM recurrent layers. Fast Training Minimal pre-processing. Outperforms on the OPPORTUNITY dataset challenge.

*(A - Accelerometer, I - Inertial, K - Kinect camera, G - Gyroscope, M - Magnetometer)

Based on the comparison, the figure shows the performance of different models.

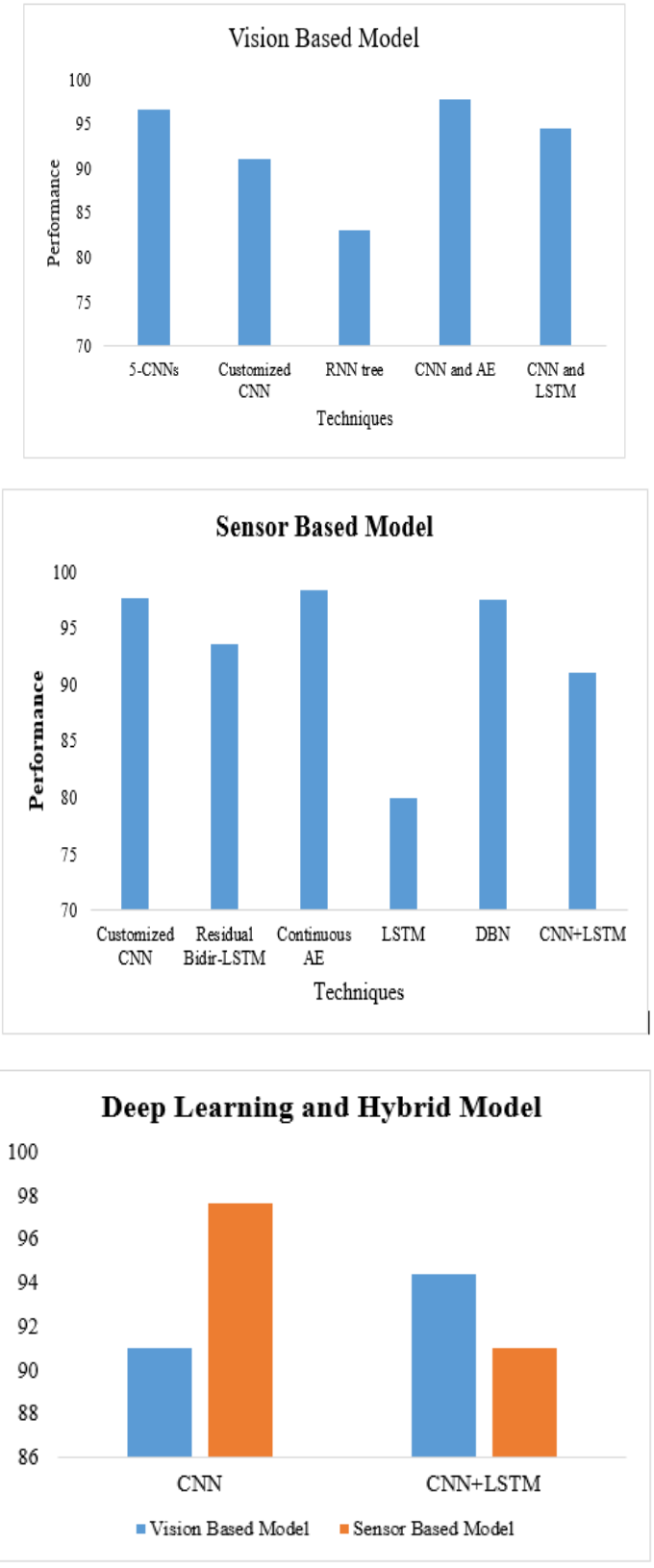

Figure 7: Performance of HAR models a) Vision HAR b) Sensor HAR c) Deep learning and Hybrid Model

4.2 Research gap

A comprehensive study of the HAR techniques shows that the existing techniques are inadequate due to the research gaps mentioned below and the need to enhance efficiency.

- Though there is a growing demand for real-time HAR applications in various fields, a research gap exists in developing efficient and accurate systems, limiting the practical implementation of HAR technologies in complex environments.
- Inter-class variability poses a major concern in many existing HAR models due to a lack of accurate classification for activities that share similarities. Developing methods capable of robustly handling these subtle differences in activities remains a significant research gap in the field.
- Integrating data from multiple sensor modalities or domains presents a research gap. HAR systems that effectively bridge Cross-Modal and Cross-Domain Recognition modalities are crucial for comprehensive activity recognition, ensuring seamless recognition across different types of data.
- Current HAR systems often focus on short-term activity recognition, leading to a lack of understanding and recognition of common long-term activities in real-life scenarios. Further exploration is needed to develop models capable of recognizing prolonged and complex activities over extended periods.
- The lack of robustness in uncontrolled environments of HAR systems, including varying lighting conditions, occlusions, and clutter, necessitates enhancements in robustness to ensure accurate recognition under diverse environmental conditions.

5. Conclusion

A comprehensive study on Human Activity Recognition is significant for providing valuable insights when selecting the proper technique from existing HAR techniques. The research extensively discovers HAR applications and the challenges faced in the field. By exploring various types of HAR techniques, including machine learning and deep learning approaches, the study sheds light on the diverse methods used for activity recognition. The research delves into both sensor-based and vision-based HAR, discussing data collection methods, sensor types, and hybrid techniques, offering a holistic understanding of the field's current landscape.

The study emphasizes the importance of addressing existing challenges, such as inter-class variability, integration of data from multiple sensor modalities, and the robustness of HAR systems in uncontrolled environments. It also highlights the limitations of deep learning and machine learning techniques, underscoring the need for further research to enhance the efficiency and accuracy of HAR models. Furthermore, the study underscores the significance of recent advancements in HAR techniques, indicating their potential to revolutionize various fields of application. The availability of openly accessible datasets and the utilization of advanced models reflect the ongoing progress in HAR research. However, the study highlights that there is a need to bridge the gaps and improve the reliability of HAR systems, ensuring their seamless integration into complex real-world scenarios.

Thus, the study not only provides valuable insights into the existing state of HAR technology but also acts as a catalyst for future research. By addressing the identified challenges and limitations, Human Activity Recognition enabled the development of more efficient, accurate, and adaptable systems for a wide range of applications.

6. Future Directions

Several fascinating potentials for future research arise from the expanding environment of Human Activity Recognition (HAR). Exploration of innovative sensor technologies, such as wearable devices and Internet of Things (IoT) sensors, is one important route for improving data gathering and activity recognition granularity. Integrating HAR with upcoming technologies like edge computing can enhance real-time, low-latency applications in a variety of applications. Furthermore, AI techniques will improve the transparency of HAR models, making them more interpretable for end-users and stakeholders. Collaborative efforts among researchers, industry experts, and policymakers are required to provide standardized datasets and evaluation standards, hence creating a more unified approach to HAR research. Also, integrating HAR with sectors such as healthcare, smart cities, and human-computer interaction will result in creative applications that will change the way we engage with technology and the environment. By embracing these future approaches, Human Activity Recognition makes significant contributions to society, enhancing the quality of life and propelling technological innovation forward.

References

[1] S. Qiu *et al.*, “Multi-sensor information fusion based on machine learning for real applications in human activity

recognition: State-of-the-art and research challenges," *Information Fusion*, vol. 80. pp. 241–265, 2022. doi: 10.1016/j.inffus.2021.11.006.

[2] Y. Liu-Thompkins, S. Okazaki, and H. Li, "Artificial empathy in marketing interactions: Bridging the human-AI gap in affective and social customer experience," *J. Acad. Mark. Sci.*, vol. 50, no. 6, pp. 1198–1218, Nov. 2022, doi: 10.1007/s11747-022-00892-5.

[3] J. Jiang, A. J. Karran, C. K. Coursaris, P. M. Léger, and J. Beringer, "A Situation Awareness Perspective on Human-AI Interaction: Tensions and Opportunities," *Int. J. Hum. Comput. Interact.*, vol. 39, no. 9, pp. 1789–1806, 2023, doi: 10.1080/10447318.2022.2093863.

[4] M. Ziaeefard and R. Bergevin, "Semantic human activity recognition: A literature review," *Pattern Recognit.*, vol. 48, no. 8, pp. 2329–2345, Aug. 2015, doi: 10.1016/j.patcog.2015.03.006.

[5] Y. Meng *et al.*, "ADAFUSE: ADAPTIVE TEMPORAL FUSION NETWORK FOR EFFICIENT ACTION RECOGNITION," in *ICLR 2021 - 9th International Conference on Learning Representations*, International Conference on Learning Representations, ICLR, 2021.

[6] J. B. Alayrac *et al.*, "Self-supervised multimodal versatile networks," in *Advances in Neural Information Processing Systems*, 2020. Accessed: Nov. 04, 2023. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/0060ef47b12160b9198302ebdb144dcf-Abstract.html

[7] J. B. Grill *et al.*, "Bootstrap your own latent a new approach to self-supervised learning," in *Advances in Neural Information Processing Systems*, 2020, pp. 21271–84. Accessed: Nov. 04, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/f3ada80d5c4ee70142b17b8192b2958e-Paper.pdf

[8] X. Liu *et al.*, "Self-Supervised Learning: Generative or Contrastive," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 1, pp. 857–876, 2023, doi: 10.1109/TKDE.2021.3090866.

[9] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in Vision: A Survey," *ACM Comput. Surv.*, vol. 54, no. 10, 2022, doi: 10.1145/3505244.

[10] P. Pareek and A. Thakkar, "A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications," *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 2259–2322, Mar. 2021, doi: 10.1007/s10462-020-09904-8.

[11] Mohan, A., Prabha, G. and V., A. 2023. Multi Sensor System and Automatic Shutters for Bridge- An Approach. International Journal of Intelligent Systems and Applications in Engineering. 11, 4s (Feb. 2023), 278–281.

[12] Prabha , G. , Mohan, A. , Kumar, R.D. and Velrajkumar, G. 2023. Computational Analogies of Polyvinyl Alcohol Fibres Processed Intellgent Systems with Ferrocement Slabs. International Journal of Intelligent Systems and Applications in Engineering. 11, 4s (Feb. 2023), 313–321.

[13] Study On Structural Behaviour Of Ductile High-Performance Concrete Under Impact And Penetration Loads, Lavanayaprabha, S. Mohan, A. Velrajkumar, G., Mohammedharoonzubair, A. Journal of Environmental Protection and Ecology., 2022, 23(6), pp. 2380–2388.

[14] Mohan, A., & K, S. . (2023). Computational Technologies in Geopolymer Concrete by Partial Replacement of C&D Waste. International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 282–292.

[15] Mohan, A., Dinesh Kumar, R. and J., S. 2023. Simulation for Modified Bitumen Incorporated with Crumb Rubber Waste for Flexible Pavement. International Journal of Intelligent Systems and Applications in Engineering. 11, 4s (Feb. 2023), 56–60.

[16] R.Gopalakrishnan, Mohan, "Characterisation on Toughness Property of Self-Compacting Fibre Reinforced Concrete", Journal of Environmental Protection and Ecology 21, No 6, 2153–2163 (2020)

[17] M. Tammvee and G. Anbarjafari, "Human activity recognition-based path planning for autonomous vehicles," *Signal, Image Video Process.*, vol. 15, no. 4, pp. 809–816, Jun. 2021, doi: 10.1007/s11760-020-01800-6.

[18] D. R. Faria, M. Vieira, C. Premebida, and U. Nunes, "Probabilistic human daily activity recognition towards robot-assisted living," in *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2015, pp. 582–587. doi: 10.1109/ROMAN.2015.7333644.

[19] A. Hayat, M. D. Fernando, B. P. Bhuyan, and R. Tomar, “Human Activity Recognition for Elderly People Using Machine and Deep Learning Approaches,” *Inf.*, vol. 13, no. 6, 2022, doi: 10.3390/info13060275.

[20] L. Schrader *et al.*, “Advanced Sensing and Human Activity Recognition in Early Intervention and Rehabilitation of Elderly People,” *J. Popul. Ageing*, vol. 13, no. 2, pp. 139–165, Jun. 2020, doi: 10.1007/s12062-020-09260-z.

[21] K. Host and M. Ivašić-Kos, “An overview of Human Action Recognition in sports based on Computer Vision,” *Heliyon*, vol. 8, no. 6. 2022. doi: 10.1016/j.heliyon.2022.e09633.

[22] S. Zhang *et al.*, “Deep Learning in Human Activity Recognition withWearable Sensors: A Review on Advances,” *Sensors*, vol. 22, no. 4, 2022, doi: 10.3390/s22041476.

[23] A. Omolaja, A. Otebolaku, and A. Alfoudi, “Context-Aware Complex Human Activity Recognition Using Hybrid Deep Learning Models,” *Appl. Sci.*, vol. 12, no. 18, 2022, doi: 10.3390/app12189305.

[24] G. Diraco, G. Rescio, A. Caroppo, A. Manni, and A. Leone, “Human Action Recognition in Smart Living Services and Applications: Context Awareness, Data Availability, Personalization, and Privacy,” *Sensors*, vol. 23, no. 13, 2023, doi: 10.3390/s23136040.

[25] L. Minh Dang, K. Min, H. Wang, M. Jalil Piran, C. Hee Lee, and H. Moon, “Sensor-based and vision-based human activity recognition: A comprehensive survey,” *Pattern Recognit.*, vol. 108, no. 05006, pp. 143–747, 2020, doi: 10.1016/j.patcog.2020.107561.

[26] A. S. Syed, D. Sierra-sosa, A. Kumar, and A. Elmaghraby, “smart cities IoT in Smart Cities : A Survey of Technologies , Practices and Challenges,” pp. 429–475, 2021.

[27] J. Park and S. Kim, “Study on Strengthening Plan of Safety Network CCTV Monitoring by Steganography and User Authentication,” vol. 2015, pp. 1–10, 2015.