 doi:10.20944/preprints202102.0349.v4

Human Activity Recognition with Deep Learning: Methods, Progress & Possibilities

Pranjal Kumar*, *DoCSE, NIT Hamirpur, pranjal@nith.ac.in;*

Abstract

Over the past decade, recognition of human activities (HAR) has become a vibrant field of research in particular, the spread in our every day lives of electronics such as mobile phones, smart cell phones and video cameras. Furthermore, the advancement in the field of deep methodologies and other paradigms have enabled scientists to enable HAR in many areas, consisting of activities in fitness and wellness. For instance, HAR is one of many resorts to support older people through day to day activities to support their cognition and physicality. This study is centered on the key aspects deep learning plays in the development of HAR applications. Although numerous HAR examination studies were carried out previously, there have been no overall studies on this subject, in all the earlier studies there were only specific HAR-related subjects. A detailed review covering all the main subjects in this area is therefore essential. This study discusses the latest developments and works in HAR. It separates the methods and the advantages and disadvantages of each method group. This paper finally addresses many problems in the current HAR subject and provides recommendations for potential study.

Learning (artificial intelligence); Neural networks; Activity recognition

I. Introduction

Human Activity Recognition (HAR) is becoming a highly researched subject over many years now, given its influence in many fields, including overall well-being, games, sport, and surveillance systems [1]. In addition, the ageing population is now becoming one of the main problems of the nation. It has been projected that by 2050, total number of old age people will rise significantly [2]. Overwhelming increment of this sort would bear many important social and health insinuations. HAR is an important tool for controlling the physicality, functionality, and cognition of older adults at home [3].

Primary goal of human activity recognition is to consider the activities in regulated and unregulated environments. Despite numerous applications, corresponding algorithms still have to come to terms with many challenges that includes: 1) intricacy and variation in day-to-day activities, 2) variability with the subject for the same operations, 3) to maintain balance in privacy and performance, 4) embedded and portable system calculation reliability, and 5) data annotation difficulty [2], [4]. Data is normally obtained from two key sources for HAR training and testing, embedded and ambient sensors. Data obtained from ambient sensor such as temperature sensors or video cameras placed in particular location in the surroundings can be ambient sensors [5], [6]. Embedded sensing is integrated into or built into clothes and other specialised medical devices, for instance smart phones and smart clocks [7]–[10]. The HAR applications are widespread in the use of cameras, but video data collection presents a wide range of privacy and computational issues [11]. Though video input generate abundant information about the context of subject/scene, the constraints in confidentiality have led the community to operate with additional environmentally friendly along with integrated sensors, including depth data. The standard workflow for developing HAR methodologies is shown in Figure 1.

HAR research has seen an major and significant expansion in the usage of Deep Learning methodologies as an algorithm of implementation, resulting in an improvement in accuracy of recognition [6], [8]. In most HAR applications as of now Classic Machine Learning (CMLs) models can be suitably adapted due to the limited data set, reduced data dimension and the accessibility of expert know-how when designing the problem [12]. In many HAR applications, DL methods yield highly accurate results with larger activity datasets. The rising interest in HAR can be linked to the increasing usage with wearable sensory data points and appliances in all aspects of everyday life, particularly for applications of health and wellness.

II. Existing Surveys

Research on recognition of human activity (HAR) has progressed considerably over the last decade. Applications that are successful include surveillance [13]., smart home [14], video analytics [15]., autopilot [16]. HAR's objective is to identify the behaviour of a user so that computer systems can proactively support the user [17]. HAR [18] consists of two main classes: vision and sensor. The vision-based HAR has produced fruitful findings with the advantage of high optical sensor resolution and the rapidly evolving computer vision techniques [19]–[22]. Since HAR is a significant subject of research in recent years, numerous investigations have been published. The burgeoning of sensors is very attached to their potential to track the orientation of the human body directly. Wang, Jindong, et al (2019) has presented the HAR state-of-the-art modalities for sensors, classification, operation, including both the traditional and newer methods, with a focus on the approach linked with each HAR stage [23]. Moreover, HAR, which includes video data and arrangement that incorporate wearable and ambient sensors, is provided. Sousa Lima et al. (2019) present a comprehensive, overview of the current techniques and solutions for smartphone based HAR whihc primarily incorporates inertial sensors and Elbasiony and Gomaa [24] (2019) include a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202102.0349.v4

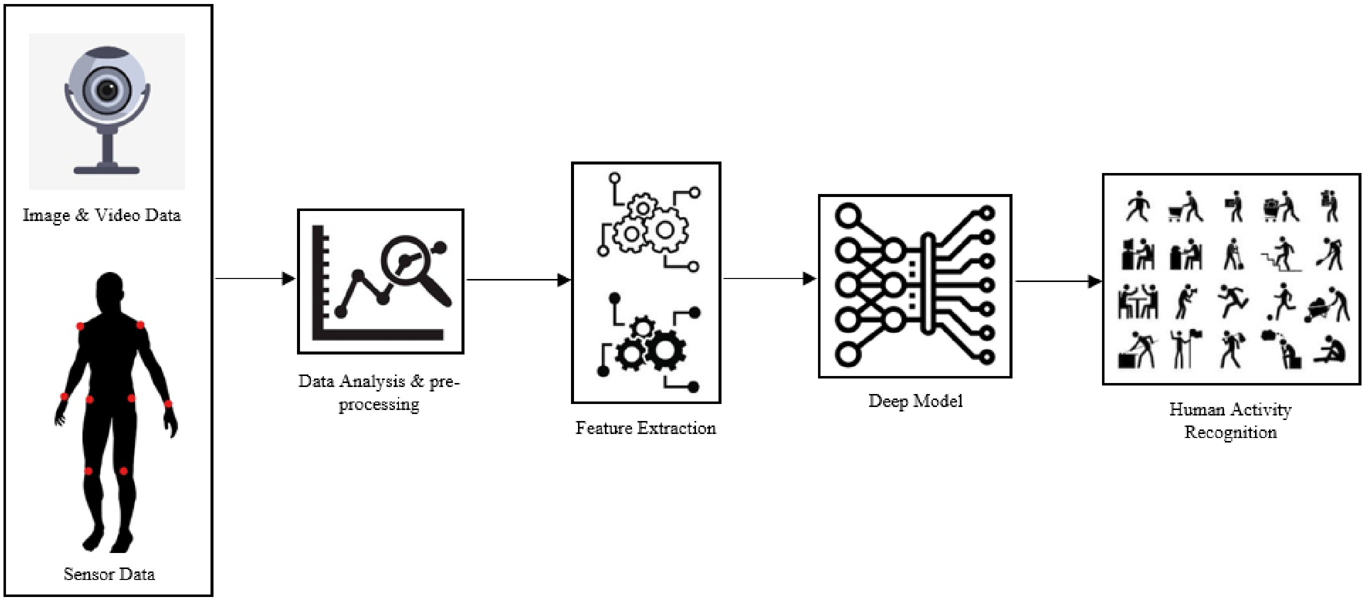

Fig. 1. The workflow of human activity recognition

thoroughness survey of several HAR systems in portable sensors (Accelerometer, Gyroscope and Magnetometer). D.Triboan et al. [25] reports an approach to fine-grained AR which combines multimodal data from individual objects and addresses the imprecise nature of non-binary sensor measurements. This approach uses fuzzy ontology to model fine-grained behaviour, with imprecise Member States of the sensors, method for classifying action completion with entity and fuzzyDL reasoning. A non-intrusive, heterogeneous ambient and embedded object-based sensing approach is also proposed for the microserver architecture. Darpan Triboan et al. provide semiotic theory based on the ontological model, capturing generic information and residential expectations for the performance of ADLs to help the segmentation process [26].

The comprehensive study of fusing data/sensors and various HAR categorization systems with focus on mobile and wearable devices is given by Nweke et al. [27] (2019). In the study of around 50 papers on physiologic data generated by sensors in medical applications including electromyography (EMG), electrocardiogram(ECG), electrooculograms (EOG), and electroencephalograms(EEG), Faust et al. [28] (2018) stated that deep learning works better than machine classification methods for large and varied data sets. A comprehensive outline of machine learning and data analytics techniques used in activity recognition, which identify with fundamental problems and challenges, was provided by Ramasamy Ramamurthy and Roy [29] (2018). Eventually, Morales et al. [30] (2017) offer a rundown of state of the art data collection, signals and pre-processing techniques along with various on-body positions and inclination calibration, choice of the right set of functions, modelling, classification models with how the HAR system can be evaluated. In addition, it includes repetitive movements, postures, falling and inactivity identification. Most of these reviews concentrate significantly on methods of data organization and model identification. [31] presents a knowledge-based approach to the continuous detection of activity based on multi-sensor data in intelligent homes. In [32] an exhausting survey is presented to investigate how different aspects of sensor activity recognition evolve and current status. Models of activities play an important role in the awareness and support of activities in an environmentally supported living environment. A variety of issues, such as cold starting, model reusability and incompleteness, affect existing methods of activity modelling. [33] implement a hybrid ontological approach to Activity Modeling in order to deal with these issues that combines domain knowledge-based modelling and data-driven model learning.

A. *Contributions*

In this work, we offer a distinctive contribution to the literature by considering a wider view of overall HAR research development over the last decade. Contrary to current surveys, we don't concentrate solely on the algorithmic information. Most of the study/works analyses/analysed only a specified aspect of the HAR as explained in the previous section. In addition, several new hypotheses, procedures and applications have recently been added by the introduction of profuse amount of HAR modelling frameworks and methodologies. Consequently, an exhaustive HAR survey is significant and pivotal for collaborators/contributor, doctors and researchers that are trying to formulate and make integration of these methods into existing systems or conduct more enhanced HAR research. This survey summarises past research and deals with a wide range of aspects of the HAR, such as data collection, methods and models for activity recognition.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202102.0349.v4

III. Background

HAR algorithms are mainly aimed at acknowledging activities based on data collected from wearables along with environmental sensors [23], [34]. These behaviours are primarily recognised on the basis of Machine Learning(ML) and Deep Learning(DL) techniques. We discuss basic ML and DL principles and methodologies in this section.

A. Context of Machine Learning

It is an artificial intelligence division that has been developed to develop algorithms which can deduce patterns from a training dataset [35]. Two main classes of these algorithms are supervised and unsupervised. The primary aim of the supervised scheme is to develop and forecast future unseen data points using a mathematical model designed on the basis on the relation between the input and the output data(initially provided to the model). The objective is to classify input-data patterns without knowing the result in unsupervised learning [5]. One or more stages of pre-processing, including the extracting of features, vectoring/segmenting, standardisation or normalisation, and projecting, are usually often essential [36].
Some of the widely popular ML controlled algorithms are: Naive Bayes, k-Means , Support Vector Machine (SVM) and Regression. By sorting them according to features/data values, DT classifies data instances. Each node is a categorised function and each branch is incidentally a value that the particular node will have to assume. Bayes' theorem with independence suppositions is founded upon application of NB classification systems. Support vector machines are based on the idea that a hyper-plane on either side will divide the two groups of data distinctively. Eventually, K-NN is aslo a prominently used ML algorithm that store and classify all provided cases by similarity (e.g. distances including Euclidean, Manhattan, Minkowski) [35]. In addition, since HAR has unique limitations like decrement in latency, memory restrictions and computing restrictions, all the above mentioned classifiers, with the exception of the SVM, are suitable to low-resource environments due to their low requirements in computer and storage.
The most widely adopted algorithms are k-mean, hierarchical clusters and hybrid models, amongst uncontrolled and especially clustering algorithms. The aim of K-Means clustering is to separate sample groups in clusters based on the similar and different data points measurements. Mixture model is defined as a probability model that reflects observer sub-populations in the total population [5]. These techniques are especially suitable for processing label-free sets of data, or if a primary outcome [37]–[39] is to quantify similarity/differences between groups.

B. Context of Deep Learning

HAR research have had a major development of Deep Learning (DL) practices in terms of algorithmic implementation, leading to increased accuracy [40]–[43] and because of its superior performance in many fields, these techniques have become extremely popular in recent years [5]. It is structured on the scheme of the depiction of data, so that optimum properties are generated automatically from raw data without human intervention, allowing the unknown pattern to be identified which would otherwise remain hidden or unknown [44]. Data sets are used in many HR applications, however, because of their small data size, reduced input data dimension, and access to expert knowledge in formulating the problem [12], Classical machine learning (CML) models may be better suited for many HAR applications. Increasing interest in HAR can be associated in various facets of daily life, particularly for physical health and well-being , with escalating use of the sensors and wearable devices. A few of the widely frequent algorithms in this domain are: Convolutionary Neural Nets, the CNN, the Long Short Memory Networks(LSTM), the Gated Recurrent Network (GRU), encoder, the TCN, and variational autoencoders [45]. CNNs are now a major tool, in particular in the field of image processing. CNN imposes local connectivity on the raw data by considering the given image as an assembly of various patches of local pixels , which extract more important features. In [46] the authors suggested that data gleans from a three-axis accelerometer would be converted into an image format and then CNN was used for identifying human activities, having three convolutionary layers and one fully connected layer. [47] proposed a classification for activity recognition which combined deep CNN and LSTM in order to classify 27 gestures of the hands and five movements. Simulation results showed that 0.93 and 0.958 were respectively the F1 score of the two classifiers. Instead of traditional micro-doppler imaging, Lin et al. [48] presented a novel iterative CNN strategy with pre-processing capability for self correlation.

C. Human Activity

Daily life activities (ADLs) are broadly defined. ADL's are all things we carry out every day, such as feeding, washing, dressing, working, homemaker, recreation and physical activity. The HAR scientific analysis offers an overview of the ADL's most researched.
Among the ADLs, to walk, to run, to stand are the most common activities in HAR study. However, in the recent years, other behaviours like various stages of cooking, [49] house-cleansing, [50]–[52], smoking [53], swimming [54] etc., were also investigated. Other activities, including complex activities, were studied in recent years. Various experiments are conducted at certain sites like to sit on the floor, to lay on the bed [55], to walk and to stand in the elevator [56], walk and run on treadmill, walk in a parking area, stepping [57] or practising on a cross-trainer [57]. Additional comprehensive recognition of movement

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202102.0349.v4

includes complex weapon movements like transportion of an object, its release, front height and various different actions to be carried out in connection with other objects [58], [59]. One main field of HAR research is population ageing and the increase in the total number of persons with physical and cognition affliction. Many activity recognition models are being deployed in order to assist individuals in identifying along with preventing risks as that of falling in older adults in parkinson's disease [60]–[62] or freezing gait (foG) [34]. In addition, ADLs are becoming common for activity tracking devices. These devices can very accurately also evaluate physical as well as physiological and markers, for example, increase/decrease in heart rate, fluctuations in blood pressure, total steps taken, shift in levels along with consumed calories. State-of-the-art instruments are able to detect sleep with sleep neurological phases [63] (i.e. nREM and REM); all of the information processed can also be used as a HAR algorithm [63].

IV. Sensor Modalities

Sensors are devices that quantify the world around us. physical aspects. The ability to gain knowledge about human activities is important. The recognition of human activity plays an important role in the daily life of people. The need to infer various easy to complex human activities is a priority to solve many human-centered problems, such as health care and individual assistance. Therefore, it is essential for the systems of human activity recognition to consistently be categorised as sensing technology. In two categories, sensor-based HAR approaches can generally be considered: data or knowledge-driven approaches. Data-led approaches use data sets to learn models of activity through the use of machines and data mining technology [64], while knowledge-driven approaches build up models by exploiting rich prior knowledge in the field of interest [65]. While some HAR approaches can be generalised to all sensor modalities, most are specific to specific types only. The modalities are mainly classified in three aspects (Chavarriaga et al. 2013), namely, body-wearing sensors, object sensors and ambient sensors [66].

A. Body Worn Sensors

One of the most common modes in HAR is body-wear sensors. Sensors such as a speedmeter, magnetometer, and gyroscope are usually worn by users. Acceleration and angular rate are modified depending on the movement of the human body, and thus human activities can be inferred. Smart phones, watches, bands, glasses and helmets are often able to find such sensors. Activity recognition in the ubicomp and wearable-computing research communities with body-worn smart devices, including smart phones, smart-watches and smart glass, has been actively studied, because activity recognition using such sensors may apply in both home sites, such as analysis of the daily lives of elderly patients and indigenous people [67]–[69].

Using the traditional PR methods, wearable sensor-based activity recognition mainly includes 3 steps, i.e. sensor data collection, extraction of features and model training [4]. First, the acquisition of sensor data such as accelerometers, gyroscopes, magnetometers, sensors for electromyography, audio sensors or vibration sensors. Secondly, the data based on human experience or field knowledge are manually taken from features such as the time domain, frequency domain, or statistical features. Finally, the characteristics of models are used to identify activities [70]. For the following reasons, deep learning models have been preferred over traditional pattern recognition (PR) [23], [71]:

- Traditional PR methods rely mainly on human experience and domain know-how [71] to extract the characteristics from heuristic and handcrafted methods.
- Traditional pattern recognition methods require a large number of labelled models data while deep networks can use unlabeled model training data.
- Traditional pattern recognition models focus on static data training, while in reality, activity data is transmitted in real time, requiring robust, incremental learning.

B. Object Sensors

The wearable and ambient sensors are used to focus people's motions. But, besides simple activities (e.g. walking, sitting, jogging and others), human beings carry out composite activities (e.g. drinking, cooking, playing) by constantly interacting in practical scenarios with the environment. Consequently, it is essential to incorporate information about use of objects to recognise more complicated human activities.

- Radio frequency identification (RFID): RFID sensors are the most widely used to identify object use with regard to cost-efficiency, reliability and easy implementation [72]–[75]. RFID tags should be attached to target objects like mugs, books, computers and toothpastes when acting as objects rather than as ambient sensors [76]. A worn RFID reader is also required in the detection phase. A sensor read is treated as a binary mark to indicate if the object is used.

C. Ambient Sensor

Environmental sensors are used to record human-environmental interactions. Usually, they are integrated into the intelligent environment of users. There are numerous types of environmental sensors like radars, sound sensors, sensors for pressure and temperatures. Ambient sensors are used for the capture of environmental change in contrast to object sensors that measure the object movements.

 doi:10.20944/preprints202102.0349.v4

Smart homes [77], equipped with sensors, can be equipped for monitoring different environmental and resident conditions [78] and drives for efficient help in their daily activities [79], for technology to be available for AAL. The design may include service or social robots, which may introduce additional functionalities and tools or provide more natural human-robot-interactions to enhance these environments and improve their acceptance towards end-users [80]. The ability to collect visual information with less privacy than with fixed cameras is a benefit of introducing these robots to an AAL environment [81]. Robot behaviours can also be used to manage situations sensitive to privacy [82].

D. *hybrid Sensors*

The use of air pressure sensor data and the inertial measuring unit for improving the HAR framework is indicated by Fu et al. [83] (IMU). This HAR model has a performance of close to 2% higher than other models that are not applicable to sensors. The HAR studies have moved from device-bound strategies into device-free approaches in the last decade. A WiFi-based HAR framework is introduced by Cui et al., using Channel State Data for the recognition of common activities. WiFi-based HAR can, however, only detect basic behaviours such as functionality and standing. It is because CSI can't understand dynamic events sufficiently well [84].

Sensor-based HAR uses data from a wide range of sensors integrated with wearable devices to capture time series [85], [86]. The authors study a context-conscious HAR system in [87] and note that an accelerometer is mainly suitable to detect simple walking and sitting activities. When gyroscope data are added, system detection performance for complex activities such as drinking, eating, and smoking is increased. However, some work is being done to explore other sensors.

V. Video Modalities

The broad spectrum of applications such as the human-computer interaction [88], [89], the analysis of sport-related performance [90], and video surveillance [91], [92] have been attracting increasing concern in recent years. While research has progressed in this area, many challenges remain, such as high changes in the shape of the human body, clothing and viewpoints, and systems acquisition conditions.

A. *Skeletal/Body-Joints*

Generative approaches are taken for evaluating/examining the location of various body parts in different frames of video [93]. These particular set of methods use beforehand information for evaluation, such as motion data [94] and contextual data [95], based on joints and rigid parts. The introduction of Toshev et al. [96] by the "DeepPose" project led HPE researchers to change from classical to profound learning approaches. ConvNets is the prominent building block that significantly take place of traditional hand-crafted features and graphical models with latest developments in pose estimation systems. For pose estimation along with action detection, Gkioxari [97] incorporated a CNN architecture.

Domain of "pose estimation" has recently been improved by Nibali et al. [98]. The thermal mapping policies which were initially adopted in 2D pose estimates have been extended to the task of 3D pose estimation. Toyoda et al. [99] presented a procedure so as to increment the input data with rotational increase. The authors then implemented the pose estimate methodology numerous number of times to single frame utilizing "post-data augmentation" techniques in order to boost the quality of exaggerated motion's pose estimations. The most stable/constant pose was then chosen and a transition for smoothing was followed.

B. *RGB and RGB-D*

Currently, an in-depth and local combination of features is popular for the recognition of activity [21], [100]. Feng et al. [101] proposed the multilayer LSTM network approach to geometrical relation features in order to detect human activity using information on the skeleton joints. Keceli et al. [102] presented an approach to the recognition of dyadic activity from the depth sequence of 3D and 2D CNN. They extract temporal features by 3D CNN trained volumes of 3D depth, while 2D CNN is adjusted to weighted sequences of sum depth. A multimodal Action Recognition model based on feature extraction and depth videos using the CNN architecture was proposed by Ijjina and Chalavadi [103]. The human activities from these fused features are recognised through an ELM classifier. The spatiotemporal hybrid neural network presented by Jing et al. [104] characterise human activities in a 2-stream neural network based RGB-D video. Srihari et al. [105] introduced a four-stream CNN network to address human activity consisting of two RGB-D video recordings and two optical temporal movement streams.

VI. Discussion & Challenges

We presented a summary of the latest HAR research in this paper. HAR is a crucial field in the identification of activity, computer technology and human support. In recent decades, HAR is becoming much more important with increasing new technologies and rising needs, including the ageing population. DL-based HAR methods have achieved outstanding recognition results in recent years. But CML-based methods are still commonly used and without any computational costs they produce excellent performance. The reproductivity of ML models has nevertheless become more and more relevant in recent years. According to our research the findings are not fully reproducible due to proprietary data sets for 78% of the proposed HAR

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202102.0349.v4

methods. This leads to obstacles for the research community to identify and benchmark the best models.
In addition, the absence of heterogeneous public data sets decreases the opportunity to create HAR models with greater capacity for generalisation. The data used in the documents examined was mainly obtained in a regulated environment. The inter and intra-subject variability that is missing in such scription data sets compounds this issue because the majority of the proposed HAR models are only evaluated and collected in a single checked setting on a small number of activities. Of the 149 HAR models analysed, close to 90 were examined with only one dataset, while over 60 were tested with several datasets. A further significant problem is the elucidated findings, prominently related to results with similar methods along with examination on the same dataset, which say that the accuracy of activity recognition results is almost identical. Such a problem involves testing using proprietary software, a shortage of open access of the code and writers that are not providing their source code publicly. Furthermore, the heterogeneity of the data and a HAR method concept, which recognises the behaviours of individuals with various motor and physical features, coincide with the data sources for data collection. A number of sensors and instruments are often used to collect data, as we have shown. In terms of data source, the methodologies suggested are nevertheless typically very rigid. Especially when using a certain sensor(s) and then changing a sensor model it becomes difficult for a specific person to test a methodology. Different sensors possess various technological features, which often include their particular phase, e.g. error in measurement or jitter presented by a particular sensing module. Classic machine learning methods are considered more familiar than complex and deep models in the case of HAR. The reason for this is that CML models need less training data and lower computer requirements. Furthermore, deep learning models are hard to comprehend inherently. Nevertheless, these models are exceptional in the knowledge and precision of more complex activities. Furthermore, a data preprocessing step is not needed. The decision of consideration of the exact deep or machine learning model is focused significantly on the computer specifications as well as the amount of annotated data availability for training.
With recognition of human activity several forms of deep learning methodologies have been used. However, many technological challenges still face this area. Many of the problems are associated with other areas of pattern recognition, including machine vision and analysis of natural languages, while others are specific to sensor-based activity recognition. Below are few examples of issues that the recognition group will tackle.

VII. Conclusion & Future Direction

In the last decade, HAR systems have become an increasing field of research and have made remarkable progress. In particular, sensor based activity recognition offers numerous precedence compared with vision based approaches that are confidentially and computationally limited. ML and DL-based activity identification algorithms become centralised in HAR. From a comprehensive examination of the current HAR studies that forms the basis of the most studied human activities we analysed the examined literature, the most widely-used electronic data source are sensors and most popular equipment amalgamate with the sensors without the requirement of video-based techniques. Details were of great interest to sensor-based data from physiological, inertial and environmental sensors. Also thoroughly studied are the product categories categorised in: a) stand-alone devices, b) smartphones or c) smartwatch devices. The results were shown for each group based on the average activities recognised, the total number of data sets incorporated to check the methods as well as the average accuracy parameter was taken into account for most works. This study do examines accelerometer, gyroscope and magnetometer-based methodologies also. We have addressed methods and findings on the basis of feature extraction, reduction of noise and standardisation. We provided a description of the most frequently used HAR recognition models. The most commonly used deep learning and machine learning methodologies for this reason and corresponding findings considering both in terms of consistency (exactitude) and quantity that signifies the total number of recognised activities, were also provided in this study. We have deduced the fact that HAR research community prefers to consider traditional machine learning models, especially due to the fact that they need less data and lower computing power than DL models. However, many dynamic behaviours were more fully recognised by the DL models. Future studies should concentrate on methodology growth and identification of more complicated activities with more sophisticated generalisation capabilities.
A few potential directions for research are provided below, based on reviewed articles. The deficiency of structured methods/models that can lead to miscellaneous activities carried out by a variety in users is among the key shortcomings of HAR algorithms. In order to solve a similar problem, transfer learning may reuse the information that is gained in one problem. For example, information gained on the basis of a certain inertial sensor located at a specific position in the body can be theoretically refurbished with a different sensory position or a divergent inertial sensor. Degree to the extent where transfer learning, in different scenarios could be beneficial, is not fully explored and must be examined further. The fusion of sensors offers also a successful course. The fusion of different sensors, in particular, could address the dependability and precision of a single or a multitude of sensor and enhance the data collected. Where data is not accurate from a single mode, the device may switch to another sensor mode to ensure robust data collection. Fine grains of activity recognition based on regular object experiences are another direction of study. This will help us to identify sub-actions and action sequence and provide downstream applications with much richer background knowledge. Sensor fusion may also be useful when several inertial sensors or close-up sensors are connected to everyday objects. We include a collection of recommendations to further advance progress in this area. First, benchmark datasets for the HAR population should be developed as a priority. In order to

 doi:10.20944/preprints202102.0349.v4

demonstrate an upgrade, new HAR models should be compared with existing HAR models in reference results. In addition, it is highly recommended that you create data sets with a sufficient number of topics and a range of activities. Recognition of fine-grained activities can also benefit from the standardisation of benchmarks. In addition to designing and improving HAR algorithms, researchers working on HAR algorithm should pay attention to hardware and device issues. Calculation of devices and memory, CPU and battery use analysis should be the primary objective in examining the balance between the resource use and the accuracy of recognition. Lastly, a broad study of the dependent positions and orientation may lead to inconsistent and non-robust downstream implementations, if not the design of position-dependent techniques.

- With the growing elderly population, Ambient Assisted Living (AAL) has been thoroughly studied by Smart Home (SH) in order to promote independent livelihood. Recognition of Human Activity (HAR) is the cornerstone of AAL systems for the detection of daily life activity (DLA) and timely support. Existing SH based AAL systems are primarily focused on fine grained activity identification (AR) and are effective in using binary sensors for daily items. By checking the expected object interactions with evidence from many heterogeneous sensor data the fined-grain AR receives limited attention.
- In the context-aware life assistive systems, activity recognition (AR) is central. The segmentation of sensor events observed when interleaved or concurrent everyday life activities (DDAs) take place in AR is a challenge. Several studies have proposed methods to separate and organise sensor observations and to recognise the basic or composite performance of generic ADLs. In semantically differentiating individual sensor events, however, little has been explored [25].
- Human Activity Recognition (HAR) systems with wearables usually rely on datasets manually annotated by human experts for precise timing of cases of relevant activity. However, in the mostly mobile scenarios of recognition of human activity, it is often very difficult to obtain data annotations like this. This also results in a degree of ambiguity on labels.This area can be explored more extensively [26].

REFERENCES

[1] Rodolfo S Antunes, Lucas A Seewald, Vinicius F Rodrigues, Cristiano A Da Costa, Luiz Gonzaga Jr, Rodrigo R Righi, Andreas Maier, Bjoern Eskofier, Malte Ollenschlaeger, Farzad Naderi, et al. A survey of sensors in healthcare workflow monitoring. *ACM Computing Surveys (CSUR)*, 51(2):1–37, 2018.

[2] Florenc Demrozi, Graziano Pravadelli, Azra Bihorac, and Parisa Rashidi. Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE Access*, 8:210816–210836, 2020.

[3] Yan Wang, Shuang Cang, and Hongnian Yu. A survey on wearable sensor modality centred human activity recognition in health care. *Expert Systems with Applications*, 137:167–190, 2019.

[4] Oscar D Lara and Miguel A Labrador. A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials*, 15(3):1192–1209, 2012.

[5] Ye Liu, Liqiang Nie, Li Liu, and David S Rosenblum. From action to activity: sensor-based activity recognition. *Neurocomputing*, 181:108–115, 2016.

[6] Ming Zeng, Le T Nguyen, Bo Yu, Ole J Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. Convolutional neural networks for human activity recognition using mobile sensors. In *6th International Conference on Mobile Computing, Applications and Services*, pages 197–205. IEEE, 2014.

[7] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM conference on embedded networked sensor systems*, pages 127–140, 2015.

[8] Eiman Kanjo, Eman MG Younis, and Chee Siang Ang. Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. *Information Fusion*, 49:46–56, 2019.

[9] Natalia Neverova, Christian Wolf, Griffin Lacey, Lex Fridman, Deepak Chandra, Brandon Barbello, and Graham Taylor. Learning human identity from motion patterns. *IEEE Access*, 4:1810–1820, 2016.

[10] Li Liu, Yuxin Peng, Ming Liu, and Zigang Huang. Sensor-based human activity recognition system with a multilayered model using time series shapelets. *Knowledge-Based Systems*, 90:138–152, 2015.

[11] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2625–2634, 2015.

[12] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. nature 521 (7553), 436-444. *Google Scholar Google Scholar Cross Ref Cross Ref*, 2015.

[13] Marco Cristani, Ramachandra Raghavendra, Alessio Del Bue, and Vittorio Murino. Human behavior analysis in video surveillance: A social signal processing perspective. *Neurocomputing*, 100:86–97, 2013.

[14] Qian Wan, Yiran Li, Changzhi Li, and Ranadip Pal. Gesture recognition for smart home applications using portable radar sensors. In *2014 36th annual international conference of the IEEE engineering in medicine and biology society*, pages 6414–6417. IEEE, 2014.

[15] Jessica JunLin Wang and Sameer Singh. Video analysis of human dynamics—a survey. *Real-time imaging*, 9(5):321–346, 2003.

[16] Pavlo Molchanov, Shalini Gupta, Kihwan Kim, and Jan Kautz. Hand gesture recognition with 3d convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 1–7, 2015.

[17] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3):1–33, 2014.

[18] Xiaoxia Huang and Mingwei Dai. Indoor device-free activity recognition based on radio signal. *IEEE Transactions on Vehicular Technology*, 66(6):5316–5329, 2016.

[19] Ahmad Jalal, Yeon-Ho Kim, Yong-Joong Kim, Shaharyar Kamal, and Daijin Kim. Robust human activity recognition from depth video using spatiotemporal multi-fused features. *Pattern recognition*, 61:295–308, 2017.

[20] Xiaodong Yang and YingLi Tian. Super normal vector for human activity recognition with depth cameras. *IEEE transactions on pattern analysis and machine intelligence*, 39(5):1028–1039, 2016.

[21] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. *arXiv preprint arXiv:1406.2199*, 2014.

[22] Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, and Luc Van Gool. Temporal segment networks: Towards good practices for deep action recognition. In *European conference on computer vision*, pages 20–36. Springer, 2016.

[23] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11, 2019.
[24] Reda Elbasiony and Walid Gomaa. A survey on human activity recognition based on temporal signals of portable inertial sensors. In *International Conference on Advanced Machine Learning Technologies and Applications*, pages 734–745. Springer, 2019.
[25] D. Triboan, L. Chen, and F. Chen. Fuzzy-based fine-grained human activity recognition within smart environments. In *2019 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pages 94–101, 2019.
[26] Darpan Triboan, Liming Chen, Feng Chen, and Zumin Wang. A semantics-based approach to sensor data segmentation in real-time activity recognition. *Future Generation Computer Systems*, 93:224–236, 2019.
[27] Henry Friday Nweke, Ying Wah Teh, Ghulam Mujtaba, and Mohammed Ali Al-Garadi. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Information Fusion*, 46:147–170, 2019.
[28] Oliver Faust, Yuki Hagiwara, Tan Jen Hong, Oh Shu Lih, and U Rajendra Acharya. Deep learning for healthcare applications based on physiological signals: A review. *Computer methods and programs in biomedicine*, 161:1–13, 2018.
[29] Sreenivasan Ramasamy Ramamurthy and Nirmalya Roy. Recent trends in machine learning for human activity recognition—a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1254, 2018.
[30] Jafet Morales and David Akopian. Physical activity recognition by smartphones, a survey. *Biocybernetics and Biomedical Engineering*, 37(3):388–400, 2017.
[31] L. Chen, C. D. Nugent, and H. Wang. A knowledge-driven approach to activity recognition in smart homes. *IEEE Transactions on Knowledge and Data Engineering*, 24(6):961–974, 2012.
[32] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu. Sensor-based activity recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6):790–808, 2012.
[33] L. Chen, C. Nugent, and G. Okeyo. An ontology-based hybrid approach to activity modeling for smart homes. *IEEE Transactions on Human-Machine Systems*, 44(1):92–105, 2014.
[34] Florenc Demrozi, Ruggero Bacchin, Stefano Tamburin, Marco Cristani, and Graziano Pravadelli. Toward a wearable system for predicting freezing of gait in people affected by parkinson's disease. *IEEE journal of biomedical and health informatics*, 24(9):2444–2451, 2019.
[35] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.
[36] Pedro Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012.
[37] Chelsea Dobbins and Reza Rawassizadeh. Towards clustering of mobile and smartwatch accelerometer data for physical activity recognition. In *Informatics*, volume 5, page 29. Multidisciplinary Digital Publishing Institute, 2018.
[38] Amari Vaughn, Paul Biocco, Yang Liu, and Mohd Anwar. Activity detection and analysis using smartphone sensors. In *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, pages 102–107. IEEE, 2018.
[39] Zahraa Said Abdallah, Mohamed Medhat Gaber, Bala Srinivasan, and Shonali Krishnaswamy. Adaptive mobile activity recognition system with evolving data streams. *Neurocomputing*, 150:304–317, 2015.
[40] Kun Xia, Jianguang Huang, and Hanyu Wang. Lstm-cnn architecture for human activity recognition. *IEEE Access*, 8:56855–56866, 2020.
[41] Kei Yaguchi, Kazukiyo Ikarigawa, Ryo Kawasaki, Wataru Miyazaki, Yuki Morikawa, Chihiro Ito, Masaki Shuzo, and Eisaku Maeda. Human activity recognition using multi-input cnn model with fft spectrograms. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, pages 364–367, 2020.
[42] Amin Ullah, Khan Muhammad, Weiping Ding, Vasile Palade, Ijaz Ul Haq, and Sung Wook Baik. Efficient activity recognition using lightweight cnn and ds-gru network for surveillance applications. *Applied Soft Computing*, 103:107102, 2021.
[43] Donghui Mao, Xinyu Lin, Yiyun Liu, Mingrui Xu, Guoxiang Wang, Jiaming Chen, and Wei Zhang. Activity recognition from skeleton and acceleration data using cnn and gcn. In *Human Activity Recognition Challenge*, pages 15–25. Springer, 2021.
[44] Benjamin Shickel, Patrick James Tighe, Azra Bihorac, and Parisa Rashidi. Deep ehr: a survey of recent advances in deep learning techniques for electronic health record (ehr) analysis. *IEEE journal of biomedical and health informatics*, 22(5):1589–1604, 2017.
[45] Jason Brownlee. *Master Machine Learning Algorithms: discover how they work and implement them from scratch*. Machine Learning Mastery, 2016.
[46] Yi Zheng, Qi Liu, Enhong Chen, Yong Ge, and J Leon Zhao. Time series classification using multi-channels deep convolutional neural networks. In *International Conference on Web-Age Information Management*, pages 298–310. Springer, 2014.
[47] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.
[48] Yier Lin, Julien Le Kernec, Shufan Yang, Francesco Fioranelli, Olivier Romain, and Zhiqin Zhao. Human activity classification with radar: Optimization and noise robustness with iterative convolutional neural networks followed with random forests. *IEEE Sensors Journal*, 18(23):9669–9681, 2018.
[49] Ferhat Attal, Samer Mohammed, Mariam Dedabrishvili, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat. Physical human activity recognition using wearable sensors. *Sensors*, 15(12):31314–31338, 2015.
[50] Mingqi Lv, Ling Chen, Tieming Chen, and Gencai Chen. Bi-view semi-supervised learning based semantic human activity recognition using accelerometers. *IEEE Transactions on Mobile Computing*, 17(9):1991–2001, 2018.
[51] Ye Liu, Liqiang Nie, Lei Han, Luming Zhang, and David S Rosenblum. Action2activity: recognizing complex activities from sensor data. *arXiv preprint arXiv:1611.01872*, 2016.
[52] Muhammad Arif and Ahmed Kattan. Physical activities monitoring using wearable acceleration sensors attached to the body. *PloS one*, 10(7):e0130851, 2015.
[53] Edwin Valarezo Añazco, Patricio Rivera Lopez, Sangmin Lee, Kyungmin Byun, and Tae-Seong Kim. Smoking activity recognition using a single wrist imu and deep learning light. In *Proceedings of the 2nd international conference on digital signal processing*, pages 48–51, 2018.
[54] Gino Brunner, Darya Melnyk, Birkir Sigfússon, and Roger Wattenhofer. Swimming style recognition and lap counting using a smartwatch and deep learning. In *Proceedings of the 23rd International Symposium on Wearable Computers*, pages 23–31, 2019.
[55] Donghui Wu, Zhelong Wang, Ye Chen, and Hongyu Zhao. Mixed-kernel based weighted extreme learning machine for inertial sensor based human activity recognition with imbalanced dataset. *Neurocomputing*, 190:35–49, 2016.
[56] Gabriele Civitarese, Riccardo Presotto, and Claudio Bettini. Context-driven active and incremental activity recognition. *arXiv preprint arXiv:1906.03033*, 2019.
[57] Lukun Wang. Recognition of human activities using continuous autoencoders with wearable sensors. *Sensors*, 16(2):189, 2016.
[58] Abdulhamit Subasi, Dalia H Dammas, Rahaf D Alghamdi, Raghad A Makawi, Eman A Albiety, Tayeb Brahimi, and Akila Sarirete. Sensor based human activity recognition using adaboost ensemble classifier. *Procedia Computer Science*, 140:104–111, 2018.
[59] Abdulhamit Subasi, Mariam Radhwan, Rabea Kurdi, and Kholoud Khateeb. Iot based mobile healthcare system for human activity recognition. In *2018 15th learning and technology conference (L&T)*, pages 29–34. IEEE, 2018.
[60] Genming Ding, Jun Tian, Jinsong Wu, Qian Zhao, and Lili Xie. Energy efficient human activity recognition using wearable sensors. In *2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pages 379–383. IEEE, 2018.
[61] Daniela Micucci, Marco Mobilio, and Paolo Napoletano. Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences*, 7(10):1101, 2017.
[62] Yiming Tian, Xitai Wang, Wei Chen, Zuojun Liu, and Lifeng Li. Adaptive multiple classifiers fusion for inertial sensor based human activity recognition. *Cluster Computing*, 22(4):8141–8154, 2019.

[63] Girija Chetty, Matthew White, and Farnaz Akther. Smart phone based data mining for human activity recognition. *Procedia Computer Science*, 46:1181–1187, 2015.

[64] Liming Chen, Jesse Hoey, Chris D Nugent, Diane J Cook, and Zhiwen Yu. Sensor-based activity recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6):790–808, 2012.

[65] Gorka Azkune, Aitor Almeida, Diego López-de Ipiña, and Liming Chen. Extending knowledge-driven activity models through data-driven learning techniques. *Expert Systems with Applications*, 42(6):3115–3128, 2015.

[66] Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, Sundara Tejaswi Digumarti, Gerhard Tröster, José del R Millán, and Daniel Roggen. The opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters*, 34(15):2033–2042, 2013.

[67] Quan Kong, Takuya Maekawa, Taiki Miyanishi, and Takayuki Suyama. Selecting home appliances with smart glass based on contextual information. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 97–108, 2016.

[68] Takuya Maekawa, Daisuke Nakai, Kazuya Ohara, and Yasuo Namioka. Toward practical factory activity recognition: unsupervised understanding of repetitive assembly work in a factory. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1088–1099, 2016.

[69] Francisco Javier Ordóñez Morales and Daniel Roggen. Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, pages 92–99, 2016.

[70] Emmanuel Munguia Tapia. *Using machine learning for real-time activity recognition and estimation of energy expenditure*. PhD thesis, Massachusetts Institute of Technology, 2008.

[71] Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu, and Yunhao Liu. Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)*, 54(4):1–40, 2021.

[72] Xinyu Li, Yanyi Zhang, Mengzhu Li, Ivan Marsic, JaeWon Yang, and Randall S Burd. Deep neural network for rfid-based activity recognition. In *Proceedings of the Eighth Wireless of the Students, by the Students, and for the Students Workshop*, pages 24–26, 2016.

[73] Yanwen Wang and Yuanqing Zheng. Modeling rfid signal reflection for contact-free activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1–22, 2018.

[74] Sungpil Woo, Jaewook Byun, Seonghoon Kim, Hoang Minh Nguyen, Janggwan Im, and Daeyoung Kim. Rnn-based personalized activity recognition in multi-person environment using rfid. In *2016 IEEE International Conference on Computer and Information Technology (CIT)*, pages 708–715. IEEE, 2016.

[75] Lina Yao, Quan Z Sheng, Xue Li, Tao Gu, Mingkui Tan, Xianzhi Wang, Sen Wang, and Wenjie Ruan. Compressive representation for device-free activity recognition with passive rfid signal strength. *IEEE Transactions on Mobile Computing*, 17(2):293–306, 2017.

[76] Xiaoyi Fan, Wei Gong, and Jiangchuan Liu. Tagfree activity identification with rfids. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):1–23, 2018.

[77] Giuseppe Amato, Davide Bacciu, Stefano Chessa, Mauro Dragone, Claudio Gallicchio, Claudio Gennaro, Hector Lozano, Alessio Micheli, Gregory MP O'Hare, Arantxa Renteria, et al. A benchmark dataset for human activity recognition and ambient assisted living. In *International Symposium on Ambient Intelligence*, pages 1–9. Springer, 2016.

[78] Ahmad Jalal, Shaharyar Kamal, and Daijin Kim. A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments. *Sensors*, 14(7):11735–11759, 2014.

[79] Yasser Ismail. *Internet of Things (IoT) for Automated and Smart Applications*. BoD–Books on Demand, 2019.

[80] Rainer Hasenauer, Carlotta Belviso, and Irmtraud Ehrenmueller. New efficiency: Introducing social assistive robots in social eldercare organizations. In *2019 IEEE International Symposium on Innovation and Entrepreneurship (TEMS-ISIE)*, pages 1–6. IEEE, 2019.

[81] Yue Yao, Jo Plested, and Tom Gedeon. Deep feature learning and visualization for eeg recording using autoencoders. In *International Conference on Neural Information Processing*, pages 554–566. Springer, 2018.

[82] Francisco Erivaldo Fernandes, Guanci Yang, Ha Manh Do, and Weihua Sheng. Detection of privacy-sensitive situations for social robots in smart homes. In *2016 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 727–732. IEEE, 2016.

[83] Zhongzheng Fu, Xinrun He, Enkai Wang, Jun Huo, Jian Huang, and Dongrui Wu. Personalized human activity recognition based on integrated wearable sensor and transfer learning. *Sensors*, 21(3):885, 2021.

[84] Zawar Hussain, Michael Sheng, and Wei Emma Zhang. Different approaches for human activity recognition: A survey. *arXiv preprint arXiv:1906.05074*, 2019.

[85] Emiro De-La-Hoz-Franco, Paola Ariza-Colpas, Javier Medina Quero, and Macarena Espinilla. Sensor-based datasets for human activity recognition–a systematic review of literature. *IEEE Access*, 6:59192–59210, 2018.

[86] Narit Hnoohom, Sakorn Mekruksavanich, and Anuchit Jitpattanakul. Human activity recognition using triaxial acceleration data from smartphone and ensemble learning. In *2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 408–412. IEEE, 2017.

[87] Sumeyye Agac, Muhammad Shoaib, and Ozlem Durmaz Incel. Context-aware and dynamically adaptable activity recognition with smart watches: A case study on smoking. *Computers & Electrical Engineering*, 90:106949, 2021.

[88] LIU Qiang, Wei Zhang, LI Hongliang, and King Ngi Ngan. Hybrid human detection and recognition in surveillance. *neurocomputing*, 194:10–23, 2016.

[89] Andrea D'Eusanio, Alessandro Simoni, Stefano Pini, Guido Borghi, Roberto Vezzani, and Rita Cucchiara. Multimodal hand gesture classification for the human–car interaction. In *Informatics*, volume 7, page 31. Multidisciplinary Digital Publishing Institute, 2020.

[90] Luis Unzueta, Jon Goenetxea, Mikel Rodriguez, and Maria Teresa Linaza. dependent 3d human body posing for sports legacy recovery from images and video. In *2014 22nd European Signal Processing Conference (EUSIPCO)*, pages 361–365. IEEE, 2014.

[91] Masoumeh Rahimi, Alireza Alghassi, Mominul Ahsan, and Julfikar Haider. Deep learning model for industrial leakage detection using acoustic emission signal. In *Informatics*, volume 7, page 49. Multidisciplinary Digital Publishing Institute, 2020.

[92] Cheng Chen, Yi Yang, Feiping Nie, and Jean-Marc Odobez. 3d human pose recovery from image by efficient visual feature selection. *Computer Vision and Image Understanding*, 115(3):290–299, 2011.

[93] Xianjie Chen and Alan Yuille. Articulated pose estimation by a graphical model with image dependent pairwise relations. *arXiv preprint arXiv:1407.3399*, 2014.

[94] Ben Daubney, David Gibson, and Neill Campbell. Estimating pose of articulated objects using low-level motion. *Computer Vision and Image Understanding*, 116(3):330–346, 2012.

[95] Huazhong Ning, Wei Xu, Yihong Gong, and Thomas Huang. Discriminative learning of visual words for 3d human pose estimation. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[96] A Toshev and C'Deeppose Szegedy. Human pose estimation via deep neural networks'. *CVPR.(Columbus, Ohio, 2014)*, pages 1653–1660.

[97] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[98] Aiden Nibali, Zhen He, Stuart Morgan, and Luke Prendergast. 3d human pose estimation with 2d marginal heatmaps. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1477–1485. IEEE, 2019.

[99] Kohei Toyoda, Michinari Kono, and Jun Rekimoto. Post-data augmentation to improve deep pose estimation of extreme and wild motions. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1570–1574. IEEE, 2019.

[100] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

 doi:10.20944/preprints202102.0349.v4

[101] Jiageng Feng, Songyang Zhang, and Jun Xiao. Explorations of skeleton features for lstm-based action recognition. *Multimedia Tools and Applications*, 78(1):591–603, 2019.

[102] Ali Seydi Keceli, Aydın Kaya, and Ahmet Burak Can. Combining 2d and 3d deep models for action recognition with depth information. *Signal, Image and Video Processing*, 12(6):1197–1205, 2018.

[103] Earnest Paul Ijjina and Krishna Mohan Chalavadi. Human action recognition in rgb-d videos using motion sequence information and deep learning. *Pattern Recognition*, 72:504–516, 2017.

[104] Chao Jing, Ping Wei, Hongbin Sun, and Nanning Zheng. Spatiotemporal neural networks for action recognition based on joint loss. *Neural Computing and Applications*, 32(9):4293–4302, 2020.

[105] D Srihari, PVV Kishore, E Kiran Kumar, D Anil Kumar, M Teja Kiran Kumar, MVD Prasad, and Ch Raghava Prasad. A four-stream convnet based on spatial and depth flow for human action classification using rgb-d data. *Multimedia Tools and Applications*, pages 1–24, 2020.