Pomiary Automatyka Robotyka, ISSN 1427-9126, R. 27, Nr 2/2023, 31–46, DOI: 10.14313/PAR_248/31

Predicting Human Activity – State of the Art

Ekemeyong Awong Lisiane Esther, Teresa Zielińska
Faculty of Power and Aeronautical Engineering, Warsaw University of Technology

Abstract: Predicting human actions is a very actual research field. Artificial intelligence methods are commonly used here. They enable early recognition and classification of human activities. Such knowledge is extremely needed in the work on robots and other interactive systems that communicate and cooperate with people. This ensures early reactions of such devices and proper planning of their future actions. However, due to the complexity of human actions, predicting them is a difficult task. In this article, we review state-of-the-art methods and summarize recent advances in predicting human activity. We focus in particular on four approaches using machine learning methods, namely methods using: artificial neural networks, support vector machines, probabilistic models and decision trees. We discuss the advantages and disadvantages of these approaches, as well as current challenges related to predicting human activity. In addition, we describe the types of sensors and data sets commonly used in research on predicting and recognizing human actions. We analyze the quality of the methods used, based on the prediction accuracy reported in scientific articles. We describe the importance of the data type and the parameters of machine learning models. Finally, we summarize the latest research trends. The article is intended to help in choosing the right method of predicting human activity, along with an indication of the tools and resources necessary to effectively achieve this goal.

Keywords: activity prediction, inferring human action, robot-human interaction

1. Introduction

People naturally are able to understand and predict to a certain extent actions of others, which is the foundation for good communication and interaction in daily life. Like for humans, for robots it is often very important to understand human activities at an early stage before they are completely executed, in order to be able to provide a timely and proper reaction [1]. Human activity recognition (HAR) can be viewed as the task of identification, and naming of human-performed activities with the help of artificial intelligence (AI) using sensory data. Understanding human activity is a very important aspect of intelligent machine vision [2]. Human activities are often influenced by natural feelings such as tiredness, exhaustion, loss of attention, low patience, need for rest, limited physical strength, etc and these result in low efficiency when it comes to performing necessary tasks. Hence, HAR is important for implementation in human assisting robots to support human tasks [3]. The examples are monitoring crime rates using HAR, HAR for elderly people care observes activity patterns and reacts in case of change of behavior or an unrecognized event. HAR applied for monitoring physical activities helps to manage and reduce the risk of contracting health problems such as obesity, diabetes, cardiovascular diseases, etc. It also helps to improve some performances for example in sports activities by detecting and estimating human poses with the arm to improve the motion dynamics. HAR also enables creation of smart home environments.

Moreover, medical diagnosis can be made using physiological measurements and observation, e.g recognizing activities like smoking, sunbathing, etc can help diagnose patients' health issues. Human activity prediction (HAP) can be referred to as a process of early inferring human activity from partial observations. HAP approaches basically aim at interpreting human actions, gestures, behaviors and correctly classifying them into respective categories before their complete execution [4].

The topics of HAP and HAR have brought a significant contribution to technological advancements. The main difference between HAR and HAP is the decision time. HAP is more important for before-the-fact quick-decision making situation rather than after-the-fact conclusion using HAR (Fig. 1). Nevertheless, both play a significant role in human-human interaction [5], human-object interaction [6], and human-machine interaction [7, 8]. This research domain has highly contributed in many areas including sports [9], robotics [10], security [11], healthcare [12]. The most common sensors used for data recording include RGBD cameras [13, 14], CCTV-came-

Autor korespondujący:
Teresa Zielińska, teresaz@meil.pw.edu.pl

Artykuł recenzowany
nadesłany 16.11.2022 r., przyjęty do druku 23.05.2023 r.

ras [15], wearable sensors [16], radio frequency identification (RFID) tags [9], gyroscopes and accelerometers [17, 18]. The problem of HAR and HAP have been challenging for recent years. However, due to the availability of low cost and low power sensors, bigger computational resources, advancements in computer vision, machine learning (ML) and AI, there is significant growth in these research areas. HAP is implemented in various environments including factory facilities allowing effective worker-robot collaboration, road navigation by self-driving vehicles, smart offices, etc. During HAP system design, it is crucial to consider factors such as the type of available devices, method of analysis, and area of application [19]. Obtaining reliable results is also highly dependent on the quality of human pose representation and activity prediction methods. Several methods have been proposed for HAP and HAR and thanks to these methods, researchers are able to explore every thinkable aspect of HAR and HAP. In this work, we will summarize the Artificial Neural Networks (ANN), Support Vector Machine (SVM) approach, probabilistic methods, and decision tree approaches. Each of the specific methods often shows varying performance depending on the application.

2. Objective of the Paper

A recent review [20] gives an elaborated overview focusing on existing deep learning-based approaches in HAR and HAP. It explains methodologies, along with related features considered, and datasets used in previous works. In this paper, we present the state-of-art methods in HAP and HAR. We review the various contributions brought by different HAP methods. The types of sensors, data types, and the corresponding advantages and challenges associated with these methods are presented. The structure of the paper is as follows: section 3 presents a review of the sensors used for HAP and HAR. The sources and types of input data, and the data acquisition techniques associated with them are described. Section 4 presents a review of the machine learning approaches used for both complex and simple activities prediction. It continues elaborate the types of interactions involved in human activities, the stages of the data processing, and the challenges of HAP and HAR works. Section 5 presents the discussion of the reviewed state of the art methods of HAP. It continues to give the overall comparison between the machine learning methods, advantages and disadvantages of these methods considering the factors such as accuracy, computing efficiency, time consumption, etc. Section 6 describes the conclusion of our work.

3. HAP/HAR Devices and Datasets

The type of devices and data used for HAP depend on the intended application. Activity data can be obtained from various types of sources, the most common are the following sources: wearable sensors, video cameras, RFID systems, and Wi-Fi devices [21]. Data can be recorded both ways: remotely (contactless) e.g radars, videos, or directly with physical contact with a person like e.g accelerometers and gyroscopes. Many years ago, sensor data collection for HAP was very expensive and challenging but with the advancement of technology, this task has become relatively easier, and several open source datasets are made available for public use.

3.1. HAP/HAR Devices

3.1.1. Vision-based Devices (Camera)

Although HAP and HAP tasks differ as illustrated in the Fig. 1, common methods may be used for sensory data collection. Vision data consist of a series of video frames as inputs and a processed output is usually a label that identifies the activity. In the paper [22], actions were recognized using static images. Many types of cameras are applied for data collection including RGB-D cameras and classical CCTV cameras. CCTV cameras observe the subject and the environment where the activity is being performed and deliver classical images. Activities are recognized using machine learning models [23, 24]. The data provided by the RGB-D camera is advantageous in the sense that it delivers several modalities such as RGB images, depth, audio, etc which improves the amount of extracted information and hence increases accuracy of the resulting predictions [18, 25]. In CCTV predictions accuracy is strongly dependent on lightning and background color, which is rather not the case in RGB-D. The depth information of RGB-D camera can be used to develop 3D skeleton-based HAP. Skeleton-based HAP and HAR use neural networks methods to

Fig. 1. HAR vs HAP taking into account picking the box activity captured in Warsaw University of Technology laboratory
Rys. 1. Interpretacja HAR i HAP dla czynności obejmującej podniesienie pudełka – zbór danych Politechniki Warszawskiej

extract the skeleton joints position in the 3D scene from recorded RGB-D images. Some works [26–28] used RGB-D data for HAP and HAR, others used only the depth data [29–31] or the skeleton data [1, 32]. Vision-based systems are advantageous in the sense that they are not invasive. It can be tedious for a disabled person or elderly to wear sensors to record data. A single RGB-D camera can be used to replace several sensors. However, cameras raise concerns about privacy issues and can be hacked [33].

3.1.2. Wearable Devices

A wearable sensor device consists of a sensory set integrated in a wearable system or can be directly attached to the body to monitor human activity, e.g inertial sensors [16, 34–38]. Due to the stressful and invasive nature of attachable sensors, wearable sensor approaches are used more often. Examples include smart phones' gyroscope and accelerometers, smart watches [17, 39–42]. Wearable sensory devices are advantageous because they offer higher privacy unlike cameras, and they can collect more accurate physiological data.

3.1.3. RFID Devices

RFID is a wireless communication method that detects a body using electromagnetic or electrostatic coupling in the radio frequency portion of the electromagnetic spectrum. It consists of an RFID tag and a reader. Data is collected by placing the RFID tag close enough to the user, and a reader collects the data when the user is in the sensing range. RFID tags can be of passive or active type. Passive type uses the energy from the reader to operate, on the other hand, active tags operate with integrated battery. Compared to passive tags, the sensing range of active tags is often larger, however passive tags are more cost-effective [24]. During sensing, multiple RFID tags are attached to the moving parts of the body. During motion, radio frequency is constantly received by the reader antennae. The difference in the timings of signals reception from various tags is used to calculate the motion of the body i.e phase difference of the radio frequency signal caused by change in relative distance between antennae and tags during movement enable tracking of the body parts in 3D scenes. In the paper [43], an adaptive generalized ANN used RFID data to track human poses in real--time. Recorded RFID data is rather prone to noise created by moving people and objects, and data processing can be a challenging task because features extraction from noisy data is often required [44]. However, some researches decide to feed RFID signals directly to the processing algorithm for HAR [45], without any filtering. RFID systems are robust to dynamic environments, which implies they are flexible and adaptable to changing environmental setting e.g warehouses, manufacturing facilities, home environments, outdoor environments, etc.

3.1.4. Wi-Fi Devices

Wi-Fi technology uses radio waves to transmit information through the air. It is often inexpensive and easy set up, which is a significant advantage. Researchers are working on using Wi-Fi devices to capture human activity data. Radio waves released by Wi-Fi transmitter are reflected by parts of the human body. Motion of humans and objects during HAR and HAP are sensed by tracking and analyzing the changes in the reflected signal caused by variation in wave propagation paths due to the movement. The data gathered using Wi-Fi transmission show high quality and are used in machine learning algorithms [46–48]. Data are organized in the form of channel state information (CSI) matrix, which holds information on properties of the communication signal such as phase shifts, power decay, etc. In the paper [49] Adaptive Activity Cutting Age Algorithm (AACA) used amplitude information from the Wi-Fi device's CSI to distinguish between body moving parts and non-moving parts during activity recognition. Results described in several other works such as [50, 51] showed the effectiveness of Wi-Fi devices in data acquisition for HAP and HAR applications. A limitation of Wi-Fi devices is their high sensitivity to dynamic environments [52].

3.2. HAP/HAR Datasets

Numerous datasets have been made publicly available for HAR and HAP research. Some examples of commonly used datasets are listed as follows;

CAD-60 and CAD-120 datasets: these datasets contain RGD-D videos of 4 people performing several activities in different settings e.g the living room, the office, and the kitchen, recorded using the Microsoft Kinect sensor. These activities are made of several atomic actions. The dataset consists of RGB images, depth images, and 3D skeleton data (x, y, z joints coordinates). CAD-60 [53] comprises of 60 RGB-D videos and CAD-120 [54] has 120 RGD videos. Some of the videos include: having a meal, drinking water, opening a pill container, stacking books, etc.

Florence 3D action dataset: this dataset from Florence University contains 9 activities recorded using Kinect camera. There are 215 RGB-D videos in this dataset, activities were performed by 10 people. Some of the activities include; "answer the phone", "read a watch", "clap", "stand up", etc. The paper [55] used human skeleton data from Florence 3D dataset to recognize human actions based on joints positions.

UFC50 dataset: machine learning methods often consider pixels from RGB videos as input data. A good example of dataset to be used that way is UFC50. It contains youtube videos grouped into 50 categories, where each category comprises of varying number of sample videos. Videos are realistic, with varying viewpoints, objects, illumination, etc, and are not staged like in the case of many other datasets [56]. Examples of activities in this dataset include; "horse riding", "basketball", "golf swing", etc.

KTH dataset: this dataset consists of grayscale videos of 25 individuals performing 6 activities in 4 different scenarios i.e "indoor", "outdoor", "outdoor with scale variation", and "outdoor with different clothes" [57]. The variation in the scenarios enable the machine learning algorithm to be tested in the real world, which has varying scenes. Activity labels include "hand clapping", "boxing", "jogging", "hand waving", "walking", "running".

Weizmann dataset: it is an action-focused dataset that consists of RGB video data grouped in 10 actions performed by 9 subjects [58]. The actions were recorded in a plain background using a fixed camera. Some of the actions include "bend", "jumping jack", "jump", etc.

UCI-HAR dataset: this dataset contains values of accelerations and angular velocities along x,y, and z axes recorded at a rate of 50Hz, using embedded accelerometer and gyroscope in a smart phone that was attached to the waist of the subject [59]. A total of 6 activities were performed by 30 people. The activities include "standing", "laying", "walking", "walking upstairs", "walking downstairs", and "sitting".

Motion Sense dataset: the motion sense dataset like the UCI-HAR dataset was recorded using the accelerometer and gyroscope embedded in a smart phone [60]. The phone was placed in the subject's front pocket to capture angular velocity and acceleration of 24 people performing 6 activities labeled "downstairs", "upstairs", "walking", "jogging", "sitting", and "standing". The dataset contains times series information of the user's altitude, yaw, pitch, row, angular velocity, and linear acceleration along the x, y and z axes for every performed activity.

WiAR dataset: this is a Wi-Fi based human activity dataset that comprises of 16 activities performed in 3 indoor environments by 10 participants, with each participant executing an activity 30 times. The activities are classified into three categories: 1) upper body e.g high throw, toss paper, 2) lower body e.g forward kick, side kick, 3) whole body activities e.g squat, sit down. The Wi-Fi data consists of the Received Signal Strength Indicator (RSSI) and Channel State Information (CSI) [61]

Wi-Fi dateset: It consists of 5 activities [62] carried out by 30 people. The activities were performed in open indoor environment as well as secluded indoor environments, since Wi-Fi signals can travel through walls. Activities include; "pick a pen from the ground", "sit down and stand up", "walking", "falling from standing position", "falling from sitting position".

The dataset in [63] consists of 245 action instances for seven different activities over 28 days. The activities are; "Leave House", "Use Toilet", "Take a Shower", "Go to Bed", "Prepare Breakfast", "Prepare Dinner", and "Get Drink". These activities were recorded using RFID technology.

Once the data collection is completed, the next step involves processing the raw sensor data to make it usable by the machine learning algorithm. This is where preprocessing comes in. Good preprocessing of raw data is compulsory for high performance in HAP and HAR tasks [64]. Preprocessing steps typically depend on the type of data collected, but can generally be classified as the following: data cleaning, normalization, transformation, feature extraction, and feature selection [65]. Data cleaning process involves fixing and regularizing the raw data i.e taking care of duplicates, inconsistent, missing data in the dataset. This is done by identifying the faults and data adjustment, updating or deleting the errors to obtain a complete data stream. In the case where video frames are used as input data, data cleaning may include resizing each frame into constant equal dimensions. Normalizing the data is the stage of preprocessing whereby the numeric values of the dataset are scaled to a common range without distorting the relative difference among them or losing the information. An example will be a Gaussian distribution technique of normalization, or normalizing the pixel values by dividing each pixel with the maximum pixel value. This simplifies the computation effort during the training and testing phases. For accelerometer data preprocessing, data segmentation is necessary because inertial data changes greatly with time. Many studies consider segments of 1 to 10 seconds [66], nevertheless the length of segments depends on the sensor sampling rate and application context. Raw data segments are later on transformed into adequate formats according to the needs of the applied machine learning methods and tools [67].

Some transformation methods include the raw plot transformation, spectrograms [68], etc. A raw plot transformation encodes inertial data e.g acceleration data into images by scaling first the data into pixels assigning the values (e.g from range 0–255), then transforming each pixel to RGB channels pixels (e.g by segmentation of normalized data into 3 groups of integers) and finally extracting the color images representing the data. After such preprocessing the data are compatible with those used by machine learning algorithms [69]. A spectrogram is a type of data representation achieved by applying the Fourier transform on the segmented data, then computing the squared magnitude to obtain a representation of the input data as a function of time and frequency. A spectrogram representation reduces variations in sensor data due to changes in the position of the sensor, change of sampling rate, altitude, etc [67]. Another common aspect of preprocessing is dimensionality reduction. This involves selection of relevant features from large datasets to facilitate detection of correlations between the features hence reducing the complexity and computation time without loss of important information. A commonly used method is Principal Component Analysis (PCA). PCA uses statistical approach by computing the eigen values and eigen vectors from the covariance matrix of the features. By ranking the eigen vectors according to their corresponding eigen values, the new features built from linear combinations of the original features can be obtained. Dimensionality reduction using laplacian eigenmaps is similar to PCA. They both emploi eigen vectors to obtain lower-dimensional data. However, PCA is a linear dimensionality reduction process, whereas Laplacian eigenmaps are based on non-linear data transformation [70].

Another approach is the Chi-square method, which uses the frequency distribution to assess the correlation between groups of features in order to select the best features. The recursive feature elimination technique proceeds by progressive elimination of features based on algorithm performance upon feeding the features, until the required number of features is obtained. Forward and backward features-selection is another method which operates by adding or discarding depending on their importance to the model performance. Feature selection algorithms are very efficient for more complex problems, capturing relevant information is not always guaranteed. Another option used by machine learning methods discussed in this work such as SVM, decision trees and probabilistic approaches commonly use hand-selected features. While hand-selected features help improve the interpretability of the method, it can be challenging in case of limited knowledge about the problem, and when dealing with complex model architectures. ANN-based methods on the other hand do not necessarily use hand-selected features or feature-selection algorithm, but automatically extract features from the input data using several layers of inter-connected nodes [71]. The final step of data preprocessing is the splitting of data into training and testing sets.

4. HAP/HAR Methods

Human activities considered in this work are divided into three main groups based on the type of interaction involved in the activity namely; human-human interaction, human-object interaction, and human-machine interaction (Fig. 2). Recent advances in HAP enabled researchers to depict people in various settings and analyze their interactions with the environment. Nevertheless, there still exist a number of challenges [72], some of which include variations in environment conditions (illumination, background), prolonged and exhaustive data collection, and proper labeling of the data in order to identify and name the raw data, based on which the machine learning model will be trained. In HAP it is relevant to consider what the person interacts with. This knowledge helps to predict the activities especially if the prediction takes into account the situation context. For example, if a person takes the cup, he or she will most likely be drinking. The various types of human interactions are further described below.

4.1. Types of Interactions

Human-human interaction: interactions between humans are usually characterized by their body positions, motion, and coordination. Human interactions can be subdivided into motion sequences, for example, a handshake interaction may consist of hand lifting, grasping, etc [21]. On the other hand, an activity may be made up of many human interactions like "greetings" activity may involve alternative interactions for example "handshake", "hug" etc. Understanding human-human interactions offer a possibility that these interactions can be predicted or even controlled. Interacting body parts are of the first importance during activity classification, therefore they are given greater weights during feature representation. Hence

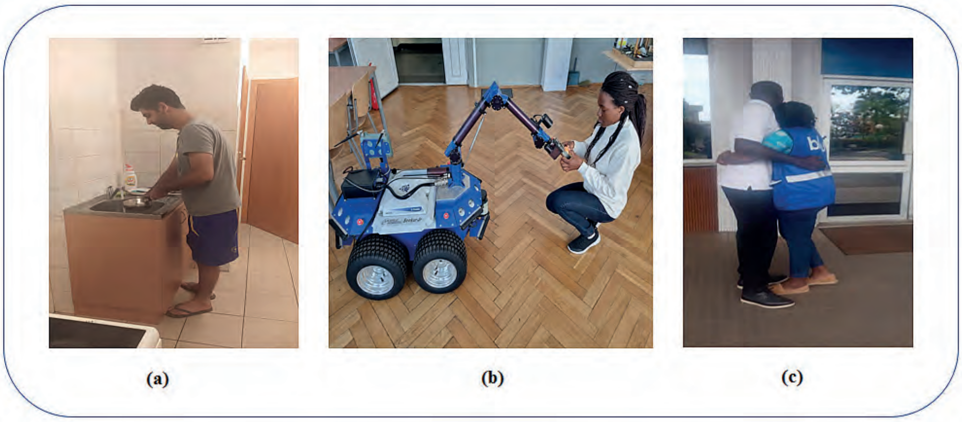

Fig. 2. Examples of interaction types considered in HAP research works. (a) human-object interaction, (b)human-machine interaction (c) human-human interaction (photo by author)
Rys. 2. Przykłady interakcji uwzględnianych w HAP: (a) interakcje człowiek-obiekt, (b) człowiek-maszyna, (c) człowiek-człowiek

paying more attention to interacting body parts improves effectiveness when capturing motion characteristics in time and space of single and simultaneously occurring activities [73]. During feature representation, body parts that interact with other objects are given more weight.

Human-object interaction: these are activities in which humans interact with objects around them. The aim is to localize and generate a relationship between them. Prediction of a human-object interaction requires knowledge about objects (e.g for what they can be used), possible interactions [74]. Xu et al. [75] proposed a scene perception architecture that predicted human tasks which involve objects. A common approach for inferring human-object interaction is by taking into account the distance between body parts and the object when learning the relationship between an object and a human [76, 77].

Human-machine interaction: these are activities in which people and automated systems interact with each other, e.g assistive robots and automated vehicles. Safety is a major issue during collaboration as the automated system has to be cautious of the end-user e.g how to navigate safely in the presence of humans, how to learn from user feedback [78]. Human-machine interaction can be remote (machine and human are separated in space and/or time), or proximate (shared human-robot environment). In the work [79] Tarik et al. Integrated HAR with robot trajectory planning so that a robot can perform the required task with high-level control by using recognized human activity as an input. It is often important that robots understand human activities in the tasks that require machine assistance, to be able to operate accordingly, hence the sensory data used in the AI algorithms must be of high quality. Figure 2 illustrates the above-discussed interactions.

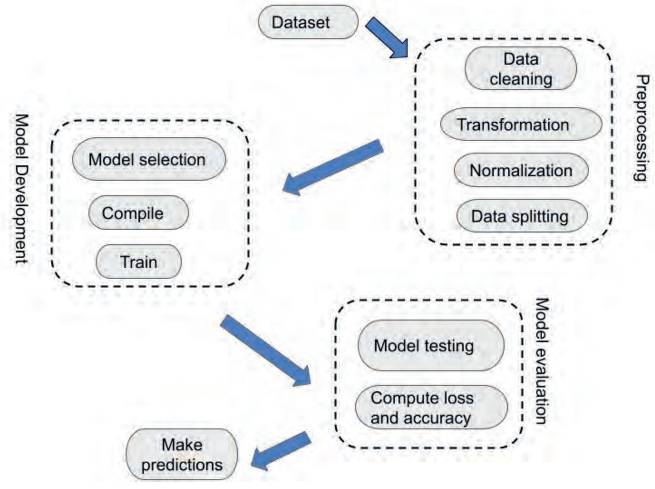

Fig. 3. Stages of data processing in HAP methods
Rys. 3. Etapy przetwarzania danych (HAP)

The HAP process consists of 4 main stages which are: 1) data acquisition and preprocessing, 2) model development, 3) model evaluation, 4) prediction. Figure 3 illustrates these stages.

4.2. The Approaches Used in HAP/HAR

Many new approaches have been tested for HAP and HAR tasks. In this paper, we reviewed 4 of those approaches proposed in recent research papers namely: Artificial Neural Networks (ANN), Support Vector Machines (SVM), probabilistic approach, and decision trees.

Artificial Neural Network(ANN): this approach attempts to emulate the human brain neurons to enable computers to learn and make decisions in a human-like way. HAP methods involving ANN with visual data require adequate human-pose representation for effective performance. The most popular representations of joint angles use Euler angles as in [80] and some use quaternions [81], whose values are fed to the learning algorithm after preprocessing. Applied to HAP, ANN learns typically by taking inputs, iteratively adjusting the weights based on the error, and supplying outputs until minimal loss is observed. Anticipation of activity at any given time is dependent not only on the present state, but also on previous observations, hence it is critical to consider the temporal aspect of activity progression [82]. The Recurrent Neural Network (RNN) is an effective model for capturing activity progression. The paper [83] demonstrated a RNN real-time surveillance system to spot violent human activities in public environments using drones. It used hand-crafted features in order to accelerate RNN learning so that it can use these features to learn more complex patterns right from the start of training. The problem with this method arises when long-term dependency is involved. To address the aforementioned limitation, RNN-based Long Short Term Memory (LSTM) was introduced [84]. LSTM consist of a memory cell that can delete and add information over time. In this way, the neural network is able to capture correlations between the previous observations as well as hidden points to provide long-term context.

Deep learning is largely applied in HAP. It uses multiple layers of ANN to progressively process the data in order to extract higher level features from combinations of lower level features. Deep learning is advantageous in the sense that it performs automatic feature extraction, which is particularly useful when dealing with challenging datasets. The paper [85], proposed a Deep Neural Network (DNN) consisting of a combination of Convolutional Neural Networks (CNN), attention layers, LSTM, and softmax layers to improve the accuracy of predictions and for short-term predicting. The attention layer selectively concentrates on the important parts of the input sequence, while the softmax is the activation function that normalizes the outputs. The paper [86] compared their

LSTM-based neural network with the Naive Bayes method. In HAP, unlike the presented LSTM approach, the Naive Bayes classification method is based on conditional probabilities for prediction of human activities and does not learn the correlations between the features. The LSTM approach reported a significantly higher prediction accuracy compared to Naive Bayes method. Deep learning is well known for the ability to auto-perform feature extraction hence this makes it suitable for HAPs using several types of input data such as images, skeleton data, time series signals, etc [87–90]. A commonly used type of neural network in HAP is the Convolutional Neural Network (CNN), due to its strong pattern detection ability especially when working with visual data. CNN contains hidden convolutional layers that process data through convolution operations using filters. The filters extract features maps detecting patterns in the data [82]. In cases of processing volumetric data or a sequence of 2D frames, the filters move through 3 dimensions of the data i.e the length, width, and height [91]. These are referred to as 3D-CNN [92]. To improve performance, 3D-CNN and 2D-CNN integration (MiCT) was proposed in the paper [93]. The spatio-temporal information for HAP was fed to 3D-CNN and 2D-CNN in parallel. Feature maps generated by each module were added together to obtain deeper feature maps for higher efficiency.

Data representation for HAP sometimes is expressed as graphs due to the highly expressive nature of graph data structure [94]. Relationships between features are then easier to understand, which allows application of Graph Neural Networks (GNN). GNN is a deep learning model that operates on graph data structure. A graph consists of nodes that interconnect through edges. Graph Convolutional Networks(GCN) is a specific type of GNN that aggregate data from neighborhood nodes in a convolutional manner. GCN uses the idea of nodes embedding. Nodes embedding involves mapping nodes to a d-dimensional embedding space, lower than the graph dimension. The nodes are mapped in such a way that the network similarity and embedded space similarity are roughly equivalent. GNN in general, has achieved high performance in HAP applications, what is advantageous is that it can perform the tasks that CNN failed to perform. Some ANN-based works include a real-time prediction of pedestrian behavior achieved in [95]. Here the mobile robots adjusted their trajectories both in time and space, and showed proactive behavior like waiting, in order to provide priority passage for pedestrians. In [96] was presented a two-stage HAR, where the first CNN discriminated between two main classes "dynamic" and "static". The second CNN was used for recognition of activities within these main classes. A drawback of neural networks is their inability to express uncertainty. To address this issue, it was proposed in [97] the Bayesian neural network to provide estimates of prediction uncertainty. It consisted of the RNN that predicted the distribution of future states of the activity, and the Bayesian network that estimated the uncertainty associated with the predictions.

Support Vector Machine (SVM): the SVM approach is a classification method. One of the most widely used in HAP is structured SVM [98]. SVMs are supervised learning methods that work by introducing a hyperplane with a maximum margin that separates data into various classes. Maximizing the margin distance enables better classification accuracy. The paper [99] presented an SVM framework that learned model parameters from a structured hierarchical representation of human activities called moveme. The SVM algorithm was trained to focus on human movements before the action was completed by cutting out the tracks in which the person makes contact e.g hug, handshake. In this way, training is done using the tracks just before contact is made therefore the algorithm focuses solely on modeling people's movements prior to action execution. The framework captured detailed characteristics of human activity that may imply a future action across frames, allowing activity prediction after observing any frame of the video, rather than using the first couple of frames as in many other works. In the paper [100], Haoran et al. suggested that performing HAP from videos didn't necessarily require using full frames from the video sample. To demonstrate this, an SVMbased early activity prediction approach was used to operate on the selected features from the keyframes of input videos. The features were compared with the activities class keyframes. Each sample video went through a key frame selection process that involved obtaining action units by tracking specific human movement across video frames. The entropy of each frame in the training video was computed, and low entropy frames were chosen as class keyframes. The structured SVM based on keyframes was developed to predict the activity class from partial observation. Some action sequences can look similar. For example, punch and push movements have many similar action sequences e.g "reach", "retract". This makes it difficult to differentiate between the action units. To address the issue Kong et al. [101] presented a structured SVM-based action prediction method that incorporated composite kernels to capture nonlinear classification boundaries when classifying the activities. The kernels transformed the training data so that the previously nonlinear classification boundaries transform into linear boundaries with higher dimensions, to facilitate linear separation of the data for more efficient activity classification.

Probabilistic Method: probabilistic methods apply statistical principles for data analysis. Although being one of the earliest forms of machine learning approaches, it is still widely used by modern researches [102]. Probabilistic approaches aim at modeling the conditional probability of variables given data distribution. One of the most popular algorithms in this class is the "Naive Bayes Classifier" [103], which is based on Bayesian reasoning. It is called "Naive" because all features are assumed to contribute independently and equally to the outcome, i.e. no pair of features is dependent. Bayesian reasoning describes the probability that a future event will occur given prior knowledge of the probability of some other event that has occurred. The Markov's model approach is another common probability-based HAP approach [104]. It assumes that at any given time in a randomly changing system, the next state is only dependent on the current state and is independent of anything in the past (1st order Markov model). An extension to the Markov model is the variable-length Markov model where the prediction of a future state depends on a fixed number of past states, with each state representing a sequence of values that may vary in length from one state to another [105]. Two mainly applied types of Markov's models include the Markov Chain and the Hidden Markov Model (HMM). Markov chains represent all system states as well as transition rates, whereas HMM is used in systems where some states are hidden, i.e. not observed. Hidden Markov Model (HMM) works well with sequential data, which is common in human activity prediction. As a result, many HMM-based methods have been proposed. A HMM-based human-machine interaction system was proposed in the work [7], whereby the user performed complex activities comprised of multiple sub-events involving various objects. The machine analyzed previously performed sub-events to predict which sub-events the user would need to complete the task, and it provided feedback to assist the user in completing the task effectively. The HMM was used to represent the sub-events in chronological order. The user's hand and the objects involved in the actions were tracked, and the relative distance between the hand and the objects were used to predict future actions.

Table 1. Advantages and disadvantages of HAP methods
Tabela 1. Zalety i wady metod HAP

Method	Advantages	Disadvantages
ANN	– Have better fault tolerance. – Network slows over learning time, and undergoes slower rate degradation. – They have numerical strength, i.e they can perform more than one job at the same time. – They can work with inadequate knowledge, i.e they may produce output even with inadequate info. It all depends on how important the missing information is. – Use of hierarchical features simplifies the learning process. – Deep learning simplicity removes the need for feature engineering which was a complex task. – Versatile nature of model make them usable for additional data, which is great for online learning' – They are reusable, e.g using an image classification for video processing, processing images for smart homes	– This network is a black box, i.e there is lack of transparency, hence difficult to trust. – Require a lot of data for training. – Poor at representing uncertainty. – Easily fooled by adversarial examples.
SVM	– Very performant where there is a clear margin of separation between classes. – More effective in high dimensional spaces. – It is relatively memory efficient.	– Comparatively less suitable for large datasets. – Do not perform very well where datasets have more noise. – Classifies data on each side of an introduced margin boundary, no probabilistic explanation for classification.
Probabilistic	– Provide an idea about the uncertainty associated with the prediction. – Ability to start off with relatively less data compared to data needed for other machine learning methods – Offer much easier explainability how the output came about.	– Time-consuming and error-prone. – Many probability methods assume predictions are independent, which hardly happens in real-world cases. – Higher complexity compared to other methods – Probability outputs estimations can be wrong in some cases.
Decision trees	– Can very well perform both regression and classification which are both relevant in HAP. – Can handle large datasets effectively, – Level of accuracy of predictions is relatively high. – Faster to train and resistant to overfitting. – Can work with missing data by creating estimates for them. – Outputs the importance of features, hence variables with positive impacts can be determined.	– Large number of trees slow down the algorithm, this can be ineffective for real-time predictions. – Require much computational power due to numerous trees. – Difficult to interpret and fails to determine the significance of each variable i.e it is predictive, not descriptive. – In cases where the predictors and outcome have a non-linear relationship, accuracy will be affected. – A small change in data may considerably change the algorithm performance

Activities belonging to the same classes considered during HAP are often regarded to occur with the same time sequence. While this is a fair assumption, in real-life applications this is not always the case. To remedy this [72] proposed an approach that handled activities with varying speed and duration using histogram of oriented velocity. Hence in this context, it is important to estimate the duration of the human activities as proposed by [106], in which a framework prediction takes into account duration of the current and future actions. The paper [6] proposed a real-time simulation for human activity identification and prediction, based on the probability of the human body intersections with the bounding colliders defined in the working space. Wearable sensors tracked body movements during human activity. When the hand intersected with the bounding collider, the action was classified using the body joints, hand positions, and corresponding bounding box. The predictions were then made by calculating the probability of the hand intersecting with a bounding box. The paper [107] used Probability Suffix Tree (PST), which is a pattern-matching technique, to model the relationship between sub-actions in an activity. The proposed approach formulated a probabilistic function that used causality reasoning and predictability parameters to predict the activity class associated with an ongoing sub-action. Each PST branch from the root to the leaves represented an activity class, with the nodes representing the probability distribution after a sub-action is executed. A Predictive Accumulative Function (PAF) generated predictability parameter which provided information on how early in the activity execution can predictions be made with satisfactory accuracy. Based on this, activities were considered as either early predictable or late predictable. PST was also used in [8] to model the relationship between actions. Each action contained information about the object, the hand position, and the type of interaction taking place. The proposed HAP framework took into account the relationships between subaction as well their temporal dependencies. First, the future action sequences were predicted using conditional probabilities, and the second part involved predicting the waiting time to execute a new action which was predicted using information of the starting time of each action, duration of each action and the sum of all intermediate action durations.

Decision Trees: decision trees are a type of non-parametric machine learning method that can be used for classification problems such as HAP [103]. It is a hierarchical system comprised of the root node, branches, internal nodes, and leaf nodes. The nodes represent the categorical features. Decision trees are applying simple decision rules based on which the tree branches are formed. Predictions are made using the majority voting rule, for multiple trees models. A decision tree-based approach to classifying human activities was proposed in the work [108]. The decision tree uses information gain to determine how relevant is a feature in determining the splitting criteria. Reduced Error Pruning (REP) was used to reduce misclassification and overfitting. REP is a bottom-up pruning approach to optimize the size of decision trees by removing non-relevant branches, which results in decreased complexity and increased accuracy.

Random Forest is a popular ensemble ML technique that combines several decision trees to improve the prediction accuracy [109]. Because HAP is a complex problem therefore learning algorithms should use a cost-effective classification learner [110]. This is offered in the Random Forest approach. The paper [111] proposed a Random Forest-based algorithm called Multi-class Balanced Random Forest (MBRF) that could predict multiple human activities simultaneously due to its ability to be scalable to multiple classes of activities. For each video frame, the approach used spatio-temporal features of defined points of interest adapted from the work [112]. To make the predictions, the similarity score was computed based on the features description. The center point location of the frame and the spatio-temporal location of the interest points with respect to the center point of the frame were taken into account. The frames with the same center point location were considered to belong to the same video. The classical support vector machine (SVM) requires solving a large quadratic programming problem with a square objective function (QPP) and linear constraints. The objective function takes into account the distance between the hyperplanes that determine the classification result. In [113], a large objective function requiring a significant computational time was divided into smaller quadratic problems according to the SMO concept. The method was combined with the Random-Forest algorithm. Minor quadratic programming problems were iteratively solved until the solution convergence criterion was met, which means obtaining a global solution. This approach was used for initial classification, and next the Random-Forest algorithm was applied. The Random Forest algorithm combines the output of multiple (randomly generated) decision trees to produce a classification score. A large number of partial decision trees are formed here. The winning classes are determined by the majority of votes. Unfortunately, the decision trees are sensitive to the data used for training. If the training data changes, the winning decision tree may be different, which means the classification result may also be different. So some change in a small amount of training data can dramatically change the forecast, even though on the scale of the full dataset such a change is marginal. One way to overcome this disadvantage is to use bootstrap aggregation (bagging), as discussed in [113]. In practice, bagging means using the average of some sets of independent data samples instead of using them independently. This reduces the dimensionality of the data and prevents overfitting of the classification, but on the other hand, it has a negative impact on the interpretability of the results.

Overall, machine learning methods for HAP and HAR are quite similar. The difference in these methods is the fact that HAR methods focus on recognising already performed tasks, while HAP focus on future tasks. In HAP, building the model considers the context of past and present activities, this can make HAP models more complex. During training, the model learns patterns and relationships among various features, and also between features and outputs. Based on these, predictions about future activities are made.

4.3. Challenges in HAP/HAR Research

As it was already mentioned HAP and HAR require data gathering, development and application of data processing method and finally testing. Therefore, we can infer that the leading challenges encountered in the HAP and HAR are associated with algorithms and datasets. Some of these challenges are listed below.

Inadequate amount of data: learning algorithms learn are using examples like a human being. But unlike the humans, large datasets are required for effective training, even for simple problems. In the paper [113] it was concluded that the significance of data surpasses the relevance of algorithms for complex problems. However, obtaining large datasets is difficult, expensive and time consuming. Nevertheless, today's researchers still obtain relatively acceptable results using small and medium datasets.

Data quality: HAP methods are not only affected by the quantity of data, but also quality of data. The work [114] researched the influence of data quality on machine learning algorithms, and it proved that data quality affects outcomes in significant ways therefore more attention should be given to data collection (experiments). It is challenging to have clean datasets without noise, errors, missing data, etc. Data therefore must be preprocessed before training. This is rather challenging task as data preprocessing methods are not predefined and depend on the nature of the problem.

Problem of overfitting and underfitting: HAP methods which use artificial intelligence and especially neural networks have chances of falling into the trap of overfitting when finding underline patterns in the data, and they cannot also find real fitting due to noisy data. This often happens especially to complex cases. This challenge can be addressed by using simpler approaches to data processing. On the other hand, underfitting is also a problem which should be approached by features processing methods.

Unnecessary features: the system needs enough amount of relevant features to perform well. Very large datasets also make it difficult to find relationships between features [115]. This is where feature engineering comes to play. Therefore dimensionality reduction algorithms like Principal Component Analysis (PCA) produce more useful but reduced sets of features [116].

Inter-subject variations and time variations: there might be a challenge when applying machine learning algorithms for predicting activities as different persons often vary in the way they perform activities. Moreover, human behaviour may also vary with time and place, i.e the same subject may perform the same activity in a slightly different manner when the environment is changed, or after a considerable period of time. These changes cause variations in datasets and hence make reasoning more challenging.

Close similarities among activities: activities that are very similar turn to have similar sensor readings. As a result, the accuracy of the predictions may be degraded in such cases.

Table 2. Recent results in HAP accuracy predictions
Tabela 2. Zalety i wady metod HAP

Article	Method	Dataset	Prediction accuracy (%)
ANN-based methods			
Sai Praneeth et al. [125]	Graph and Hierarchical Temporal Networks	CAD-120	88.9
Sadegh [126]	Multi-stage LSTM	UCF-101 UT-Interaction	80.5 84.0
Fiora Pirri et al. [127]	3D-CNN + ProtoNet + RNN	MPII Cooking	92.8
Md. ZiaUddin [128]	RNN	MHEALTH Dataset	99.69
Shi et al. [129]	Feature Mapping RNN + RBF + GAN	JHMDB-21	73.4
Neziha Jaouedi et al. [130]	RNN + CNN	CAD-60	95.5
Yaxiang Fan et al. [131]	Deep ConvNets	Hockey fight	96.9
SVM-based methods			
Yu Kong et al. [132]	Structured Support Vector Machine (SSVM)	UT1 1	86.67
M. Hoai et al. [98]	SVM	Weizmann	82
Ahmad Jalal et al. [133]	SVM	WISDM	82.77
Probabilistic methods			
J. Bütepage, et al. [134]	Probabilistic semi-supervised variational recurrent neural network (SVRNN)	UTKinect-Action3D	84.0
Siyuan Qi et al. [135]	Spatial-Temporal And-Or Graph (ST-AOG) – Probabilistic	CAD-120	86.7
Y Jin, et al. [136]	Markov Logic Network (MLN)	CAD-120	0.83
Victoria Manousaki et al. [137]	Probabilistic graphical model	CAD-120	55.9
Sheng Li et al. [138]	Dynamic Marked Point Process (DMP) + Diction of Partial Match (PPM)	MSR 3D Action Pair	69
Kang Li et al. [104]	Probabilistic Suffix tree (PST)	MPII-Cooking	79
Decision tree-based methods			
Gang Yu et al. [111]	Random Forest	UT-Interaction	90
Sheikh Badar et al. [113]	Random Forest	IMSB	81.25
Katherine Ellis [139]	Random Forest	Accelerometer data	92.3
Veralia Gabriela et al. [140]	Decision tree	Activities of Daily Living (ADLs)	88.02

5. Discussion

Recent advancements in machine learning methods for HAR and HAP have brought questions regarding which methods are more efficient, cost-effective, and time-saving. There is no direct answer to this question as every method is unique in its own way. Based on the various research works reviewed in this paper, it is clear that the performance of the HAP and HAR experiments vary due to the methodologies used, the dataset, and activity representations. Thus all of the techniques discussed have advantages and disadvantages. In the paper [117] authors aimed at finding the best machine learning algorithm for human activity prediction considering Random Forest, ANN, SVM, Naive Bayes and decision trees. According to their results, the Random Forest approach offered the highest prediction accuracy. However one can argue that the accuracy can be affected by changing the values of parameters, dataset sampling method, features selection and dimensionality reduction method, which can lead to another method showing the best accuracy. Table 1 describes the advantages and disadvantages of the discussed methods. The goal of all of the preceding techniques is to predict the activities prior to completion, so they are sensitive to prior knowledge. Deep learning is one of the most recent trends in HAP. Deep learning typically outperforms other traditional ML techniques due to its ability, unlike other methods, to extract features from raw data. However, it has a high computational cost, which is currently a rather serious limitation.

It is worth noting that the aforementioned HAP approaches are based on low-level feature representations. Researchers have found that some type of high-level representations of human complex activities called "semantic features" can be used in HAP and HAR. Semantic features are attributes that describe distinctive characteristics of the activity at a higher level. Applying these attributes improves the accuracy of predictions, particularly for activities belonging to the same class, but consisting of visually different actions caused by its execution variations [119]. While low-level features are useful for HAPs, prediction algorithms cannot anticipate "untrained" actions i.e the algorithms cannot predict an activity not seen before. To address this issue, Cheng et al. [120] proposed to manually design semantic features that represent human acti-

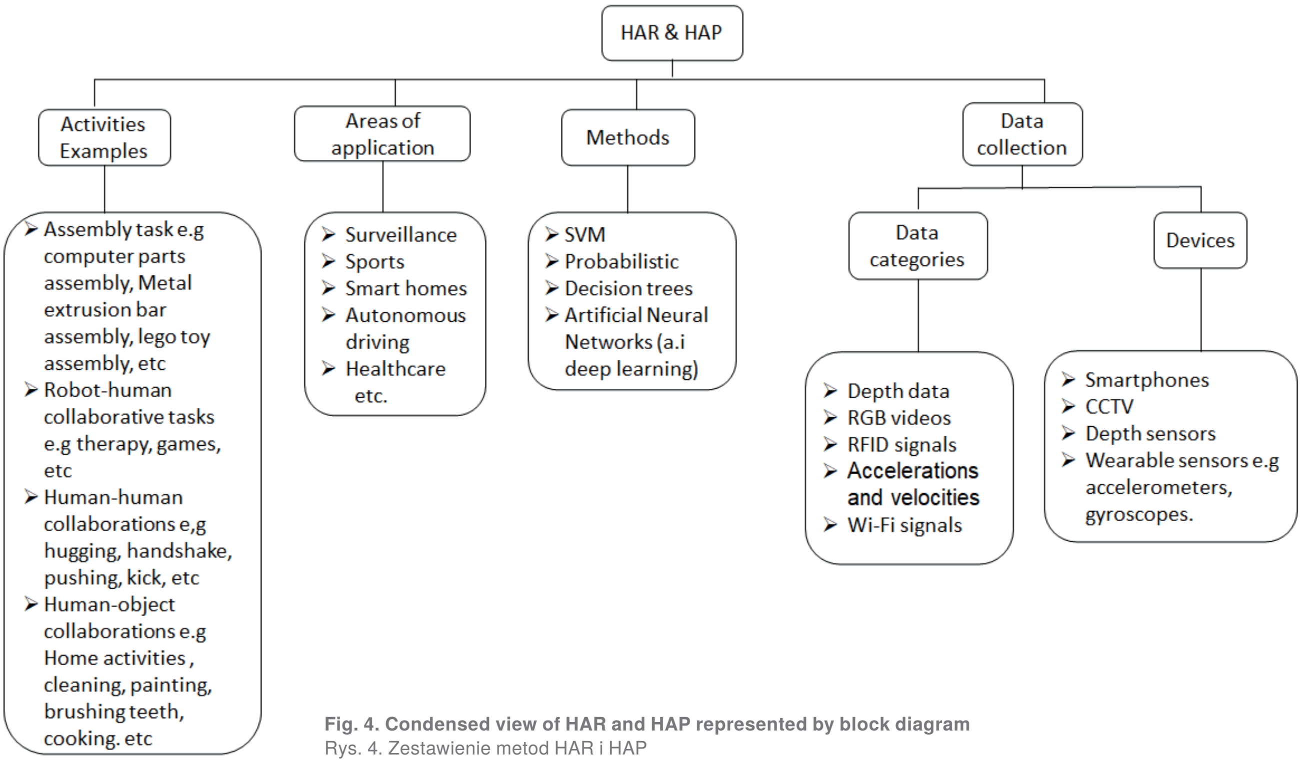

Fig. 4. Condensed view of HAR and HAP represented by block diagram
Rys. 4. Zestawienie metod HAR i HAP

vities on a higher concept. During training, each activity is characterized by the absence or presence of semantic attributes, and a classification algorithm is used to recognize the labeled characteristics corresponding to each activity. However, establishing the semantic attributes manually would be arduous, expensive, and time-consuming. To solve this problem, [121] suggested an approach based on deep learning techniques known as restricted Boltzmann machines (RBMs) and Deep Belief Networks (DBNs) to assign semantic attributes. Semantic attributes obtained by this technique showed high relevance when compared to manually assigned ones. Other methods used for capturing semantic features, include Beta-Bernoulli Process Restricted Boltzmann Machines [122], Diffusion Maps embedding [123], Markov Semantic Model [124], etc. For the view of the contents of this paper, figure 4 gives the systematic view of HAR and HAP issues.

HAP and HAR are active research and development fields which have a vast area of potential directions in the future. To mention a few; it could be focused on improving the accuracy and reliability of already existing methods of HAP and HAR be it by developing new algorithms, using better features extraction techniques, etc. Further, it could be focused for improved integration with other technologies e.g integrating sensors, cameras and other devices for higher performance and comprehensive activity tracking. Another potential future focus of HAP and HAR research could be on increasing real-world applications such as healthcare, sports, transport. With the increase in accuracy and reliability of the HAP and HAR technologies, it is likely to be used in various real-world applications.

6. Conclusion

While human activity recognition yields many promising results and is much more explored by researchers, human activity prediction is a growing area with promising potential. We reviewed 4 common types of activity prediction methods namely; ANN, SVM, probabilistic method, and decision trees. We shortly described sensor devices and datasets as well. This gives a general picture of the state of the art. It is observed that researchers have tried various techniques or approaches for better prediction of human activities using various sources of data like RGB images, depth images, RFID signals, Wi-Fi signals and data delivered by wearable sensors such as accelerations, velocities and orientation angles. Wi-Fi and RFID sensors have comparatively fewer applications due to their high flexibility to interferences. The wearable sensors on the other hand fail to sense the state of motionless objects. Hence the more common sensing technique use visual sensors, mostly RGB data. The recent advances and new ideas were summarized. Analyzing the state of the art methods of HAP enables a better understanding of steps involved in them, hence increasing the possibility of identifying efficient techniques to achieve needed performance. Table 2 gives the list of the most representative methods with their prediction accuracy. According to this table, ANN-based prediction can be concluded as the most popular and outperforming the other techniques.

References

1. Chiu H.-K., Adeli E., Wang B., Huang D.-A., Niebles J.C., *Action-Agnostic Human Pose Forecasting*, [In:] 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2019, 1423–1432, DOI: 10.1109/WACV.2019.00156.
2. Coppola C., Faria D.R., Nunes U., Bellotto N., *Social Activity Recognition Based on Probabilistic Merging of Skeleton Features with Proximity Priors from RGB-D Data*, [In:] 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2016, 5055–5061, DOI: 10.1109/IROS.2016.7759742.

3. Ravichandar H.C., Trombetta D., Dani A.P., *Human Intention-Driven Learning Control for Trajectory Synchronization in Human-Robot Collaborative Tasks*, „IFAC-PapersOnLine”, Vol. 51,No. 34, 2019, 1–7 DOI: 10.1016/j.ifacol.2019.01.001.
4. Razali H., Mordan T., Alahi A., *Pedestrian Intention Prediction: A Convolutional Bottom-Up Multi-Task Approach*, „Transportation Research Part C: Emerging Technologies”, Vol. 130, 2021, 103259, DOI: 10.1016/j.trc.2021.103259.
5. Bibi S., Anjum N., Amjad T., McRobbie G., Ramzan N., *Human Interaction Anticipation by Combining Deep Features and Transformed Optical Flow Components*, „IEEE Access”, Vol. 8, 2020, 137646–137657, DOI: 10.1109/ACCESS.2020.3012557.
6. Manns M., Tuli T.B., Schreiber F., *Identifying Human Intention During Assembly Operations using Wearable Motion Capturing Systems Including Eye Focus*, „Procedia CIRP”, Vol. 104, 2021, 924–929, DOI: 10.1016/j.procir.2021.11.155.
7. Ryoo M.S., Grauman K., Aggarwal J.K., *A task-Driven Intelligent Workspace System to Provide Guidance Feedback*, „Computer Vision and Image Understanding”, Vol. 114, No. 5, 2010, 520–534, DOI: 10.1016/j.cviu.2009.12.009.
8. Casalino A., Massarenti N., Zanchettin A.M., Rocco P., *Predicting the Human Behaviour in Human-Robot Co-Assemblies: an Approach Based on Suffix Trees*, [In:] 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2020, 11108–11114, DOI: 10.1109/IROS45743.2020.9341301.
9. Ding H., Shangguan L., Yang Z., Han J., Zhou Z., Yang P., Xi W., Zhao J., *FEMO: A Platform for Free-Weight Exercise Monitoring with RFIDs*, [In:] Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, 2015, 141–154, DOI: 10.1145/2809695.2809708.
10. Kulic D., Croft E.A., *Affective State Estimation for Human-Robot Interaction*, „IEEE Transactions on Robotics”, Vol. 23, No. 5, 2007, 991–1000 DOI: 10.1109/TRO.2007.904899.
11. Vaniya S.M., Bharathi B., *Exploring Object Segmentation Methods in Visual Surveillance for Human Activity Recognition*, [In:] 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), IEEE, 2016, 520–525, DOI: 10.1109/ICGTSPICC.2016.7955356.
12. Abobakr A., Hossny M., Nahavandi S., *A Skeleton-Free Fall Detection System from Depth Images using Random Decision Forest*, „IEEE Systems Journal”, Vol. 12, No. 3, 2017, 2994–3005, DOI: 10.1109/JSYST.2017.2780260.
13. Bandi C., Thomas U., *Skeleton-Based Action Recognition for Human-Robot Interaction using Self-Attention Mechanism*, [In:] 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), IEEE, 2021, DOI: 10.1109/FG52635.2021.9666948.
14. Li B., Tian J., Zhang Z., Feng H., Li X., *Multitask Non-Autoregressive Model for Human Motion Prediction*, „IEEE Transactions on Image Processing”, Vol. 30, 2020, 2562–2574 DOI: 10.1109/TIP.2020.3038362.
15. Li M., Chen S., Chen X., Zhang Y., Wang Y., Tian Q., *Symbiotic Graph Neural Networks for 3D Skeleton-Based Human Action Recognition and Motion Prediction*, „IEEE Transactions on Pattern Analysis and Machine Intelligence”, Vol. 44, No. 6, 2022, 3316–3333, DOI: 10.1109/TPAMI.2021.3053765.
16. Pham C., Nguyen-Thai S., Tran-Quang H., Tran S., Vu H., Tran T.-H., Le T.-L., *SensCapsNet: Deep Neural Network for Non-Obtrusive Sensing Based Human Activity Recognition*, „IEEE Access”, Vol. 8, 2020, 86934–86946, DOI: 10.1109/ACCESS.2020.2991731.
17. Zhu R., Xiao Z., Li Y., Yang M., Tan Y., Zhou L., Lin S., Wen H., *Efficient Human Activity Recognition Solving the Confusing Activities via Deep Ensemble Learning*, „IEEE Access, Vol. 7, 2019, 75490–75499, DOI: 10.1109/ACCESS.2019.2922104.
18. Wang K., He J., Zhang L., *Attention-Based Convolutional Neural Network for Weakly Labeled Human Activities' Recognition with Wearable Sensors*, „IEEE Sensors Journal”, Vol. 19, No. 17, 2019, 7598–7604, DOI: 10.1109/JSEN.2019.2917225.
19. Gupta N., Gupta S.K., Pathak R.K., Jain V., Rashidi P., Suri J.S., *Human Activity Recognition in Artificial Intelligence Framework: A Narrative Review*, „Artificial Intelligence Review”, Vol. 55, 2022, 4755–4808, DOI: 10.1007/s10462-021-10116-x.
20. Abdussami S., Nagendraprasad S., Shivarajakumara K., Singh S., Thyagarajamurthy A., *A Review on Action Recognition and Action Prediction of Human(s) using Deep Learning Approaches*, „International Journal of Computer Applications”, Vol. 177, No. 20, 2019, DOI: 10.5120/ijca2019919605.
21. Razin Y.S., Pluckter K., Ueda J., Feigh K., *Predicting Task Intent from Surface Electromyography using Layered Hidden Markov Models*, „IEEE Robotics and Automation Letters”, Vol. 2, No. 2, 2017, 1180–1185, DOI: 10.1109/LRA.2017.2662741.
22. Yao B., Fei-Fei L., *Action Recognition with Exemplar Based 2.5D Graph Matching*, [In:] European Conference on Computer Vision, Springer, 2012, 173–186, DOI: 10.1007/978-3-642-33765-9_13.
23. Diba A., Fayyaz M., Sharma V., Paluri M., Gall J., Stiefelhagen R., Gool L.V., *Large Scale Holistic Video Understanding*, [In:] European Conference on Computer Vision, Springer, 2020, 593–610, DOI: 10.48550/arXiv.1904.11451.
24. Du Y., Lim Y., Tan Y., *A Novel Human Activity Recognition and Prediction in Smart Home Based on Interaction*, „Sensors”, Vol. 19, No. 20, 2019, DOI: 10.3390/s19204474.
25. Phyo C.N., Zin T.T., Tin P., *Deep Learning for Recognizing Human Activities using Motions of Skeletal Joints,* „IEEE Transactions on Consumer Electronics”, Vo. 65, No. 2, 2019, 243–252, DOI: 10.1109/TCE.2019.2908986.
26. Mici L., Parisi G.I., Wermter S., *Recognition and Prediction of Human-Object Interactions with a Self-Organizing Architecture*, [In:] 2018 International Joint Conference on Neural Networks (IJCNN), IEEE, 2018, DOI: 10.1109/IJCNN.2018.8489178.
27. Schydlo P., Rakovic M., Jamone L., Santos-Victor J., *Anticipation in Human-Robot Cooperation: A Recurrent Neural Network Approach for Multiple Action Sequences Prediction*, [In:] 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018, DOI: 10.1109/ICRA.2018.8460924.
28. Dixon S., Hansen R., Deneke W., *Probabilistic Grammar Induction for Long Term Human Activity Parsing*, [In:] 2019 International Conference on Computational Science and Computational Intelligence (CSCI), IEEE, 2019, 306–311, DOI: 10.1109/CSCI49370.2019.00061.
29. Duckworth P., Hogg D.C., Cohn A.G., *Unsupervised Human Activity Analysis for Intelligent Mobile Robots*, „Artificial Intelligence”, Vol. 270, 2019, 67–92, DOI: 10.1016/j.artint.2018.12.005.
30. Fang L., Liu X., Liu L., Xu H., Kang W., *JGR-P2O: Joint Graph Reasoning Based Pixel-to-Offset Prediction Network for 3D Hand Pose Estimation from a Single Depth Image*,

[In:] European Conference on Computer Vision, Springer, 2020, 120–137, DOI: 10.1007/978-3-030-58539-6_8.

31. Hu J.-F., Zheng W.-S., Ma L., Wang G., Lai J., Zhang J., *Early Action Prediction by Soft Regression*, „IEEE Transactions on Pattern Analysis and Machine Intelligence", Vol. 41, No. 11, 2018, 2568–2583, DOI: 10.1109/TPAMI.2018.2863279.
32. Li M., Chen S., Liu Z., Zhang Z., Xie L., Tian Q., Zhang Y., *Skeleton Graph Scattering Networks for 3D Skeleton-Based Human Motion Prediction*, [In:] Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, 854–864, DOI: 10.1109/ICCVW54120.2021.00101.
33. Ranasinghe S., Al Machot F., Mayr H.C., *A Review on Applications of Activity Recognition Systems with Regard to Performance and Evaluation*, „International Journal of Distributed Sensor Networks", Vol. 12, No. 8, 2016, DOI: 10.1177/1550147716665520.
34. Xia K., Huang J., Wang H., *LSTM-CNN Architecture for Human Activity Recognition*, „IEEE Access", Vol. 8, 2020, 56855–56866, DOI: 10.1109/ACCESS.2020.2982225.
35. Saha J., Ghosh D., Chowdhury C., Bandyopadhyay S., *Smart Handheld Based Human Activity recognition using Multiple Instance Multiple Label Learning*, „Wireless Personal Communications", Vol. 117, No. 2, 2021, 923–943, DOI: 10.1007/s11277-020-07903-0.
36. Zhou X., Liang W., Kevin I., Wang K., Wang H., Yang L.T., Jin Q., *Deep-Learning-Enhanced Human Activity Recognition for Internet of Healthcare Things*, „IEEE Internet of Things Journal", Vol. 7, No. 7, 2020, 6429–6438, DOI: 10.1109/JIOT.2020.2985082.
37. Chen J., Sun Y., Sun S., *Improving Human Activity Recognition Performance by Data Fusion and Feature Engineering*, „Sensors", Vol. 21, No. 3, 2021, DOI: 10.3390/s21030692.
38. Tian Y., Zhang J., Chen L., Geng Y., Wang X., *Single Wearable Accelerometer-Based Human Activity Recognition via Kernel Discriminant Analysis and QPSO-KELM Classifier*, „IEEE Access", Vol. 7, 2019, 109216–109227, DOI: 10.1109/ACCESS.2019.2933852.
39. Yao S., Zhao Y., Zhang A., Hu S., Shao H., Zhang C., Su L., Abdelzaher T., *Deep Learning for the Internet of Things*, „Computer", Vol. 51, No. 5, 2018, 32–41, DOI: 10.1109/MC.2018.2381131.
40. Garcia-Gonzalez D., Rivero D., Fernandez-Blanco E., Luaces M.R., *A Public Domain Dataset for Real-Life Human Activity Recognition using Smartphone Sensors*, „Sensors", Vol. 20, No. 8, 2020, DOI: 10.3390/s20082200.
41. Lawal I.A., Bano S., *Deep Human Activity Recognition using Wearable Sensors*, [In:] Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments, 2019, 45–48, DOI: 10.1145/3316782.3321538.
42. Bashar S.K., Al Fahim A., Chon K.H., *Smartphone Based Human Activity Recognition with Feature Selection and Dense Neural Network*, [In:] 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2020, 5888–5891, DOI: 10.1109/EMBC44109.2020.9176239.
43. Yang C., Wang X., Mao S., *RFID-Based 3D Human Pose Tracking: A Subject Generalization Approach*, „Digital Communications and Networks", Vol. 8, No. 3, 2022, 278–288, DOI: 10.1016/j.dcan.2021.09.002.
44. Rohei M.S., Salwana E., Shah N.B.A.K., Kakar A.S., *Design and Testing of an Epidermal RFID Mechanism in a Smart Indoor Human Tracking System*, „IEEE Sensors Journal", Vol. 21, No. 4, 2021, 5476–5486. DOI: 10.1109/JSEN.2020.3036233.
45. Li X., Zhang Y., Marsic I., Sarcevic A., Burd R.S., *Deep Learning for RFID-Based Activity Recognition*, [In:] Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM, 2016, pp. 164–175, DOI: 10.1145/2994551.2994569.
46. Lin G., Jiang W., Xu S., Zhou X., Guo X., Zhu Y., He X., *Human Activity Recognition using Smartphones With WiFi Signals*, „IEEE Transactions on Human-Machine Systems", 2022, 1–12, DOI: 10.1109/THMS.2022.3188726.
47. Yang J., Chen X., Wang D., Zou H., Lu C.X., Sun S., Xie L., *Deep Learning and Its Applications to WiFi Human Sensing: A Benchmark and A Tutorial*, arXiv preprint arXiv:2207.07859 (2022).
48. Mei Y., Jiang T., Ding X., Zhong Y., Zhang S., Liu Y., *WiWave: WiFi-based Human Activity Recognition using the Wavelet Integrated CNN*, [In:] 2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops), IEEE, 2021, 100–105.
49. Yan H., Zhang Y., Wang Y., Xu K., *WiAct: A Passive WiFi-Based Human Activity Recognition System*, „IEEE Sensors Journal", Vol. 20, No. 1, 2019, 296–305, DOI: 10.1109/JSEN.2019.2938245.
50. Fei H., Xiao F., Han J., Huang H., Sun L., *Multi-Variations Activity Based Gaits Recognition using Commodity WiFi*, „IEEE Transactions on Vehicular Technology", Vol. 69, No. 2, 2020, 2263–2273, DOI: 10.1109/TVT.2019.2962803.
51. Ding X., Jiang T., Zhong Y., Wu S., Yang J., Zeng J., *Wi-Fi-Based Location-Independent Human Activity Recognition with Attention Mechanism Enhanced Method*, „Electronics", Vol. 11, No. 4, 2022, DOI: 10.3390/electronics11040642.
52. Li J., Jiang T., Yu J., Ding X., Zhong Y., Liu Y., *An WiFi-Based Human Activity Recognition System Under Multi-source Interference*, [In:] International Conference in Communications, Signal Processing, and Systems, Springer, 2022, 937–944, DOI: 10.1007/978-981-19-0390-8_118.
53. Sung J., Ponce C., Selman B., Saxena A., *Unstructured Human Activity Detection from RGBD Images*, [In:] 2012 IEEE International Conference on Robotics and Automation, IEEE, 842–849, DOI: 10.1109/ICRA.2012.6224591.
54. Koppula H., Saxena A., *Learning Spatio-Temporal Structure from RGB-D Videos for Human Activity Detection and Anticipation*, [In:] International Conference on Machine Learning, PMLR, 2013, 792–800.
55. Seidenari L., Varano V., Berretti S., Bimbo A., Pala P., *Recognizing Actions from Depth Cameras as Weakly aligned multi-part bag-of-Poses*, [In:] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013, 479–485, DOI: 10.1109/CVPRW.2013.77.
56. Reddy K.K., Shah M., *Recognizing 50 Human Action Categories of Web Videos*, „Machine Vision and Applications", Vol. 24, No. 5, 2013, 971–981, DOI: 10.1007/s00138-012-0450-4.
57. Schuldt C., Laptev I., Caputo B., *Recognizing Human Actions: a Local SVM Approach*, [In:] Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., Vol. 3, IEEE, 2004, 32–36, DOI: 10.1109/ICPR.2004.1334462.
58. Blank M., Gorelick L., Shechtman E., Irani M., Basri R., *Actions as Space-Time Shapes*, [In:] Tenth IEEE Interna-

tional Conference on Computer Vision (ICCV'05), Vol. 1, Vol. 2, IEEE, 2005, 1395–1402, DOI: 10.1109/ICPR.2004.1334462.

59. Anguita D., Ghio A., Oneto L., Parra Perez X., Reyes Ortiz J.L., *A Public Domain Dataset for Human Activity Recognition using Smartphones*, [In:] Proceedings of the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine learning, 2013, 437–442.
60. Malekzadeh M., Clegg R.G., Cavallaro A., Haddadi H., *Mobile Sensor Data Anonymization*, [In:] Proceedings of the International Conference on Internet of Things Design and Implementation, 2019, 49–58, DOI: 10.1145/3302505.3310068.
61. Guo L., Wang L., Lin C., Liu J., Lu B., Fang J., Liu Z., Shan Z., Yang J., Guo S., *Wiar: A Public Dataset for Wifi-Based Activity Recognition*, „IEEE Access", Vol. 7, 2019, 154935–154945, DOI: 10.1109/ACCESS.2019.2947024.
62. Baha'A A., Almazari M.M., Alazrai R., Daoud M.I., *A Dataset for Wi-Fi-Based Human Activity Recognition in Line-Of-Sight and Non-Line-Of-Sight Indoor Environments*, „Data in Brief", Vol. 33, 2020, DOI: 10.1016/j.dib.2020.106534.
63. Van Kasteren T., Noulas A., Englebienne G., Kröse B., *Accurate Activity Recognition in a Home Setting*, [In:] Proceedings of the 10th International Conference on Ubiquitous Computing, 2008, DOI: 10.1145/1409635.1409637.
64. Zheng X., Wang M., Ordieres-Meré J., C*omparison of Data Preprocessing Approaches for Applying Deep Learning to Human Activity Recognition in the Context of Industry 4.0*, „Sensors", Vol. 18, No. 7, 2018, DOI: 10.3390/s18072146.
65. Kotsiantis S.B., Kanellopoulos D., Pintelas P.E., *Data Preprocessing for Supervised Learning*, „International Journal of Computer Science", Vol. 1, No. 2, 2006, 111–117.
66. Preece S.J., Goulermas J.Y., Kenney L.P., Howard D., *A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities from Accelerometer Data*, „IEEE Transactions on Biomedical Engineering", Vol. 56, No. 3, 2008, 871–879, DOI: 10.1109/TBME.2008.2006190.
67. Ravi D., Wong C., Lo B., Yang G.-Z., *A DeepLlearning Approach to On-Node Sensor Data Analytics for Mobile or Wearable Devices*, „IEEE Journal of Biomedical and Health Informatics", Vol. 21, No. 1, 2016, 56–64, DOI: 10.1109/JBHI.2016.2633287.
68. Liu Z., Xu L., Jia Y., Guo S., *Human Activity Recognition Based on Deep Learning with Multi-Spectrogram*, [In:] 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP), IEEE, 2020, 11–15, DOI: 10.1109/ICSIP49896.2020.9339335.
69. Hur T., Bang J., Huynh-The T., Lee J., Kim J.-I., Lee S., *Iss2Image: A Novel Signal-Encoding Technique for CNN-Based Human Activity Recognition*, „Sensors", Vol. 18, No. 11, 2018, DOI: 10.3390/s18113910.
70. Bloom V., Argyriou V., Makris D., *Linear Latent Low Dimensional Space for Online Early Action Recognition and Prediction*, „Pattern Recognition", Vol. 72, 2017, 532–547, DOI: 10.1016/j.patcog.2017.07.003.
71. Khaire U.M., Dhanalakshmi R., *Stability of feature selection algorithm: A review*, „Journal of King Saud University – Computer and Information Sciences", Vol. 34, No. 4, 2022, 1060–1073, DOI: 10.1016/j.jksuci.2019.06.012.
72. Li K., Fu Y., *ARMA-HMM: A New Approach for Early Recognition of Human Activity*, [In:] Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), IEEE, 2012, 1779–1782.
73. Lee D.-G., Lee S.-W., *Human Interaction Recognition Framework Based on Interacting Body Part Attention*, „Pattern Recognition", Vol. 128, 2022, DOI: 10.1016/j.patcog.2022.108645.
74. Kambara M., Sugiura K., *Relational Future Captioning Model for Explaining Likely Collisions in Daily Tasks*, arXiv preprint arXiv:2207.09083, 2022.
75. Xu Z., Qing L., Miao J., *Activity Auto-Completion: Predicting Human Activities from Partial Videos*, [In:] Proceedings of the IEEE International Conference on Computer Vision, 3191–3199, DOI: 10.1109/ICCV.2015.365.
76. Meng M., Drira H., Daoudi M., Boonaert J., *Human-Object Interaction Recognition by Learning the Distances Between the Object and the Skeleton Joints*, [In:] 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Vol. 7, IEEE, 2015, DOI: 10.1109/FG.2015.7284883.
77. Dutta V., Zielińska T., *Predicting Human Actions Taking Into Account Object Affordances*, „Journal of Intelligent & Robotic Systems", Vol. 93, No. 3, 2019, 745–761, DOI: 10.1007/s10846-018-0815-7.
78. Uzunovic T., Golubovic E., Tucakovic Z., Acikmese Y., Sabanovic A., *Task-Based Control and Human Activity Recognition for Human-Robot Collaboration*, [In:] IECON 2018 – 44th Annual Conference of the IEEE Industrial Electronics Society, 2018, 5110–5115, DOI: 10.1109/IECON.2018.8591206.
79. Uzunovic T., Golubovic E., Tucakovic Z., Acikmese Y., Sabanovic A., *Task-Based Control and Human Activity Recognition for Human-Robot Collaboration*, [In:] IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society, IEEE, 2018, 5110–5115, DOI: 10.1109/IECON.2018.8591206.
80. Zheng X., Chen X., Lu X., *A Joint Relationship Aware Neural Network for Single-Image 3D Human Pose Estimation*, „IEEE Transactions on Image Processing", Vol. 29, 2020, 4747–4758, DOI: 10.1109/TIP.2020.2972104.
81. Pavllo D., Feichtenhofer C., Auli M., Grangier D., *Modeling Human Motion with Quaternion-Based Neural Networks*, „International Journal of Computer Vision", Vol. 128, No. 4, 2020, 855–872, DOI: 10.1007/s11263-019-01245-6.
82. Kratzer P., Midlagajni N.B., Toussaint M., Mainprice J., *Anticipating Human Intention for Full-Body Motion Prediction in Object Grasping and Placing Tasks,* [In:] 2020 29th IEEE International Conference on Robot and Human Interactive Communication (ROMAN), IEEE, 2020, 1157–1163.
83. Singh A., Patil D., Omkar S., *Eye in the Sky: Real-Time Drone Surveillance System (DSS) for Violent Individuals Identification using Scatternet Hybrid Deep Learning Network*, [In:] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, 1629–1637, DOI: 10.1109/CVPRW.2018.00214.
84. Zhao X., Chen Y., Guo J., Zhao D., *A spatial-Temporal Attention Model for Human Trajectory Prediction*, „IEEE CAA Journal of Automatica Sinica", Vol. 7, No. 4, 2020, 965–974, DOI: 10.1109/JAS.2020.1003228.
85. Putra P.U., Shima K., Shimatani K., *A Deep Neural Network Model for Multi-View Human Activity Recognition*, „PloS One", Vol. 17, No. 1, 2022, DOI: 10.1371/journal.pone.0262181.
86. Du Y., Lim Y., Tan Y., *Activity Prediction using LSTM in Smart Home*, [In:] 2019 IEEE 8th Global Conference

on Consumer Electronics (GCCE), IEEE, 2019, 918–919, DOI: 10.1109/GCCE46687.2019.9015492.

87. Zhang Q., Wang T., Wu H.-N., Li M., Zhu J., Snoussi H., *Human Action Prediction Based on Skeleton Data*, [In:] 2020 39th Chinese Control Conference (CCC), IEEE, 2020, 6608–6612, DOI: 10.23919/CCC50068.2020.9189122.
88. Dong M., Xu C., *Skeleton-Based Human Motion Prediction With Privileged Supervision*, „IEEE Transactions on Neural Networks and Learning Systems", 2022, DOI: 10.1109/TNNLS.2022.3166861.
89. Fragkiadaki K., Levine S., Felsen P., Malik J., *Recurrent Network Models for Human Dynamics*, [In:] Proceedings of the IEEE International Conference on Computer Vision, 2015, 4346–4354.
90. Feichtenhofer C., Fan H., Malik J., He K., *SlowFast Networks for Video Recognition*, [In:] Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
91. Ji S., Xu W., Yang M., Yu K., *3D Convolutional Neural Networks for Human Action Recognition*, „IEEE Transactions on Pattern Analysis and Machine Intelligence", Vol. 35, No. 1, 2013, 221–231, DOI: 10.1109/TPAMI.2012.59.
92. Alfaifi R., Artoli A.M., *Human Action Prediction with 3D-CNN*, „SN Computer Science", Vol. 1, 2020, DOI: 10.1007/s42979-020-00293-x.
93. Zhou Y., Sun X., Zha Z.-J., Zeng W., *MiCT: Mixed 3D/2D Convolutional Tube for Human Action Recognition*, [In:] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 449–458, DOI: 10.1109/CVPR.2018.00054.
94. Gao C., Wang T., Zhang M., Zhu A., Shi P., Snoussi H., *3D Human Motion Prediction Based on Graph Convolution Network and Transformer*, [In:] 2021 China Automation Congress (CAC), IEEE, 2021, 2957–2962, DOI: 10.1109/CAC53003.2021.9728062.
95. Schaefer S., Leung K., Ivanovic B., Pavone M., *Leveraging Neural Network Gradients within Trajectory Optimization for Proactive Human-Robot Interactions*, [In:] 2021 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2021, 9673–9679, DOI: 10.1109/ICRA48506.2021.9561443.
96. Cho H., Yoon S.M., *Divide and Conquer-Based 1D CNN Human Activity Recognition using Test Data Sharpening*, „Sensors", Vol. 18, No. 4, 2018, DOI: 10.3390/s18041055.
97. Bhattacharyya A., Fritz M., Schiele B., *Long-Term on--Board Prediction of People in Traffic Scenes Under Uncertainty*, [In:] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 4194–4202, DOI: 10.1109/CVPR.2018.00441.
98. Hoai M., De la Torre F., *Max-margin early event detectors*, „International Journal of Computer Vision", Vol. 107, No. 2, 2014, 191–202, DOI: 10.1007/s11263-013-0683-3.
99. Lan T., Chen T.-C., Savarese S., *A Hierarchical Representation for Future Action Prediction*, [In:] European Conference on Computer Vision, Springer, 2014, 689–704, DOI: 10.1007/978-3-319-10578-9_45.
100. Wang H., Yuan C., Shen J., Yang W., Ling H., *Action Unit Detection and Key Frame Selection for Human Activity Prediction*, „Neurocomputing", Vol. 318, 2018, 109–119, DOI: 10.1016/j.neucom.2018.08.037.
101. Kong Y., Kit D., Fu Y., *A Discriminative Model with Multiple Temporal Scales for Action Prediction*, [In:] European Conference on Computer Vision, Springer, 2014, 596–611, DOI: 10.1007/978-3-319-10602-1_39.
102. Ryoo M.S., *Human Activity Prediction: Early Recognition of Ongoing Activities from Streaming Videos*, [In:] 2011 International Conference on Computer Vision, IEEE, 2011, 1036–1043, DOI: 10.1109/ICCV.2011.6126349.
103. Alvee B.I., Tisha S.N., Chakrabarty A., *Application of Machine Learning Classifiers for Predicting Human Activity*, [In:] 2021 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), 2021, 39–44, DOI: 10.1109/IAICT52856.2021.9532572.
104. Li K., Fu Y., *Prediction of Human Activity by Discovering Temporal Sequence Patterns*, „IEEE Transactions on Pattern Analysis and Machine Intelligence", Vol. 36, No. 8, 2014, 1644–1657, DOI: 10.1109/TPAMI.2013.2297321.
105. Galata A., Johnson N., Hogg D., *Learning Variable--Length Markov Models of Behavior*, „Computer Vision and Image Understanding", Vol. 81, No. 3, 2001, 398–413, DOI: 10.1006/cviu.2000.0894.
106. Cheng Y., Tomizuka M., *Long-Term Trajectory Prediction of the Human Hand and Duration Estimation of the Human Action*, „IEEE Robotics and Automation Letters", Vol. 7, No. 1, 2021, 247–254, DOI: 10.1109/LRA.2021.3124524.
107. Manju D, Radha V., *An Enhanced Human Activity Prediction at Rainy and Night Times*, „International Journal of Recent Technology and Engineering, Vol. 8, 2019, 965–970, DOI: 10.35940/ijrte.C4113.098319.
108. Li C., *Wearable Computing: Accelerometer-Based Human Activity Classification using Decision Tree*, Utah State University, 2017.
109. Wang A., Chen H., Zheng C., Zhao L., Liu J., Wang L., *Evaluation of Random Forest for Complex Human Activity Recognition using Wearable Sensors*, [In:] 2020 International Conference on Networking and Network Applications (NaNA), IEEE, 2020, 310–315, DOI: 10.1109/NaNA51271.2020.00060.
110. Minor B.D., Doppa J.R., Cook D.J., *Learning Activity Predictors from Sensor Data: Algorithms, Evaluation, and Applications*, „IEEE Transactions on Knowledge and Data Engineering", Vol. 29, No. 12, 2017, 2744–2757, DOI: 10.1109/TKDE.2017.2750669.
111. Yu G., Yuan J., Liu Z., *Predicting Human Activities using Spatio-Temporal Structure of Interest Points*, [In:] Proceedings of the 20th ACM International Conference on Multimedia, 2012, 1049–1052, DOI: 10.1145/2393347.2396380.
112. Yu G., Goussies N.A., Yuan J., Liu Z., *Fast Action Detection via Discriminative Random Forest Voting and Top-K Subvolume Search*, „IEEE Transactions on Multimedia", Vol. 13, No. 3, 2011, 507–517, DOI: 10.1109/TMM.2011.2128301.
113. S. B. ud din Tahir, Jalal A., Batool M., *Wearable Sensors for Activity Analysis using SMO-based Random Forest Over Smart Home and Sports Datasets*, [In:] 2020 3rd International Conference on Advancements in Computational Sciences (ICACS), IEEE, 2020, DOI: 10.1109/ICACS47775.2020.9055944.
114. Halevy A., Norvig P., Pereira F., *The Unreasonable Effectiveness of Data*, „IEEE Intelligent Systems", Vol. 24, No. 2, 2009, 8–12, DOI: 10.1109/MIS.2009.36.
115. Sessions V., Valtorta M., *The Effects of Data Quality on Machine Learning Algorithms*, Proceedings of the 11th International Conference on Information Quality (ICIQ), Vol. 6, 2006, 485–498.
116. Chen Y.-L., Wu X., Li T., Cheng J., Ou Y., Xu M., *Dimensionality Reduction of Data Sequences for Human Activity Recognition*, „Neurocomputing", Vol. 210, 2016, 294–302, DOI: 10.1016/j.neucom.2015.11.126.

117. Ray S., Alshouiliy K., Agrawal D.P., *Dimensionality Reduction for Human Activity Recognition Using Google Colab*, „Information", Vol. 12, No. 1, 2020, DOI: 10.3390/info12010006.
118. Darji N.R., Ajila S.A., *Increasing Prediction Accuracy for Human Activity Recognition using Optimized Hyperparameters*, [In:] 2020 IEEE International Conference on Big Data (Big Data), IEEE, 2020, 2472–2481, DOI: 10.1109/BigData50022.2020.9378000.
119. Ziaeefard M., Bergevin R., *Semantic Human Activity Recognition: A Literature Review*, „Pattern Recognition", Vol. 48, No. 8, 2015, 2329–2345, DOI: 10.1016/j.patcog.2015.03.006.
120. Cheng H.-T., Sun F.-T., Griss M., Davis P., Li J., You D., *NuActiv: Recognizing Unseen New Activities using Semantic Attribute-Based Learning*, [In:] Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services, 2013, 361–374, DOI: 10.1145/2462456.2464438.
121. Venkatachalam S., Nair H., Zeng M., Tan C.S., Mengshoel O.J., Shen J.P., *Sem-Net: Learning Semantic Attributes for Human Activity Recognition with Deep Belief Networks*, „Frontiers in Big Data", 2022, DOI: 10.3389/fdata.2022.879389.
122. Mittelman R., Lee H., Kuipers B., Savarese S., *Weakly Supervised Learning of Mid-Level Features with Beta-Bernoulli Process Restricted Boltzmann Machines*, [In:] 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, 476–483, DOI: 10.1109/CVPR.2013.68.
123. Liu J., Yang Y., Saleemi I., Shah M., *Learning Semantic Features for Action Recognition via Diffusion Maps*, „Computer Vision and Image Understanding", Vol. 116, No. 3, 2012, 361–377, DOI: 10.1016/j.cviu.2011.08.010.
124. Liu J., Wang X., Li T., Yang J., *Spatio-Temporal Semantic Features for Human Action Recognition*, „KSII Transactions on Internet and Information Systems", Vol. 6, No. 10, 2012, 2632–2649, DOI: 10.3837/tiis.2012.10.011.
125. Sunkesula S.P.R., Dabral R., Ramakrishnan G., *Lighten: Learning Interactions with Graph and Hierarchical Temporal Networks for HOI in Videos,* [In:] Proceedings of the 28th ACM International Conference on Multimedia, 2020, 691–699, DOI: 10.1145/3394171.3413778.
126. M. Sadegh Aliakbarian, F. Sadat Saleh, M. Salzmann, B. Fernando, L. Petersson, L. Andersson, *Encouraging LSTMs to Anticipate Actions Very Early*, [In:] Proceedings of the IEEE International Conference on Computer Vision, 2017, 280–289.
127. Pirri F., Mauro L., Alati E., Ntouskos V., Izadpanahkakhk M., Omrani E., *Anticipation and Next Action Forecasting in Video: an End-to-End Model with Memory*, arXiv preprint arXiv:1901.03728, 2019.
128. Uddin M.Z., *A Wearable Sensor-Based Activity Prediction System to Facilitate Edge Computing in Smart Healthcare System*, „Journal of Parallel and Distributed Computing", Vol. 123, 2019, 46–53, DOI: 10.1016/j.jpdc.2018.08.010.
129. Shi Y., Fernando B., Hartley R., *Action Anticipation with RBF kernelized Feature Mapping RNN*, [In:] Proceedings of the European Conference on Computer Vision (ECCV), 2018, 301–317.
130. Jaouedi N., Perales F.J., Buades J.M., Boujnah N., Bouhlel M.S., *Prediction of Human Activities Based on a New Structure of Skeleton Features and Deep Learning Model*, „Sensors", Vol. 20, No. 17, 2020, DOI: 10.3390/s20174944.
131. Fan Y., Wen G., Li D., Qiu S., Levine M.D., *Early Event Detection Based on Dynamic Images of Surveillance Videos*, „Journal of Visual Communication and Image Representation", Vol. 51, 2018, 70–75, DOI: 10.1016/j.jvcir.2018.01.002.
132. Kong Y., Fu Y., *Max-Margin Action Prediction Machine*, „IEEE Transactions on Pattern Analysis and Machine Intelligence", Vol. 38, No. 9, 2016, 1844–1858, DOI: 10.1109/TPAMI.2015.2491928.
133. Jalal A., Quaid M.A.K., Hasan A.S., *Wearable Sensor-Based Human Behavior Understanding and Recognition in Daily Life for Smart Environments*, [In:] 2018 International Conference on Frontiers of Information Technology (FIT), 2018, 105–110, DOI: 10.1109/FIT.2018.00026.
134. Bütepage J., Kjellström H., Kragic D., *A Probabilistic Semi-Supervised Approach to Multi-Task Human Activity Modeling*, arXiv preprint arXiv:1809.08875, 2018.
135. Qi S., Huang S., Wei P., Zhu S.-C., *Predicting Human Activities using Stochastic Grammar*, [In:] Proceedings of the IEEE International Conference on Computer Vision, 2017, 1164–1172.
136. Jin Y., Zhu L., Mu Y., *Complex Video Action Reasoning via Learnable Markov Logic Network*, [In:] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, 3242–3251.
137. Manousaki V., Papoutsakis K., Argyros A., *Graphing the Future: Activity and Next Active Object Prediction using Graph-based Activity Representations*, arXiv preprint arXiv:2209.05194, 2022.
138. Li S., Li K., Fu Y., *Early Recognition of 3D Human Actions*, „ACM Transactions on Multimedia Computing, Communications, and Applications", Vol. 14, No. 1s, 2018, DOI: 10.1145/3131344.
139. Ellis K., Kerr J., Godbole S., Lanckriet G., Wing D., Marshall S., *A Random Forest Classifier for the Prediction of Energy Expenditure and Type of Physical Activity from Wrist and Hip Accelerometers*, „Physiological Measurement", Vol. 35, No. 11, 2014, DOI: 10.1088/0967-3334/35/11/2191
140. Sánchez V.G., Skeie N.-O., *Decision trees for human activity recognition modelling in smart house environments*, „Simulation Notes Europe", Vol. 28, 2018, 177–184, DOI: 10.11128/sne.28.tn.10447.

Przewidywanie aktywności człowieka – stan wiedzy

Streszczenie: Przewidywanie działań człowieka to bardzo aktualny kierunek badań. Wykorzystywane są tu powszechnie metody sztucznej inteligencji. Umożliwiają one wczesne rozpoznawanie i klasyfikowanie działań człowieka. Taka wiedza jest niezwykle potrzebna w pracach nad robotami i innymi interaktywnymi systemami komunikującymi się i współpracującymi z ludźmi. Zapewnia to wczesne reakcje takich urządzeń i odpowiednie planowanie ich przyszłych działań. Jednak ze względu na złożoność działań człowieka ich przewidywanie jest trudnym zadaniem. W tym artykule dokonujemy przeglądu najnowocześniejszych metod i podsumowujemy ostatnie postępy w zakresie przewidywania aktywności człowieka. Skupiamy się szczególnie na czterech podejściach wykorzystujących metody uczenia maszynowego, a mianowicie na metodach wykorzystujących: sztuczne sieci neuronowe, metody wektorów nośnych, modele probabilistyczne oraz drzewa decyzyjne. Omawiamy zalety i wady tych podejść, a także aktualne wyzwania związane z zagadnieniami przewidywania aktywności człowieka. Ponadto opisujemy rodzaje czujników i zbiory danych powszechnie stosowane w badaniach dotyczących przewidywania i rozpoznawania działań człowieka. Analizujemy jakość stosowanych metod w oparciu o dokładność przewidywania raportowaną w artykułach naukowych. Opisujemy znaczenie rodzaju danych oraz parametrów modeli uczenia maszynowego. Na koniec podsumowujemy najnowsze trendy badawcze. Artykuł ma za zadanie pomóc przy wyborze właściwej metody przewidywania aktywności człowieka, wraz ze wskazaniem narzędzi i zasobów niezbędnych do efektywnego osiągnięcia tego celu.

Słowa kluczowe: przewidywanie działań, przewidywanie akcji człowieka, interakcje człowiek-robot

prof. Teresa Zielińska, PhD, DSc
teresaz@meil.pw.edu.pl
ORCID: 0000-0002-8299-0727

She received her MSc degree in automatic control and PhD degree in robotics from Warsaw University of Technology (WUT), Poland. In 1995 she obtained habilitation from the Polish Academy of Sciences. She is a full professor at the Faculty of Power and Aerospace Engineering, WUT. She was the vice-chair of the Polish Branch of the IEEE Robotics and Automation Society (2016–2018), and the Secretary General of IFToMM Federation for Promotion of Machines and Mechanism Systems (2011–2019). In the years 2018–2022 she was the member of the Council of the National Science Centre. She is the local chair of Japan-Europe Master on Advanced Robotics – JEMARO. She published over 300 publications and holds 2 patents. Her research interest includes robotics and machine learning. Teresa Zielińska is the member of Committee for Automatic Control and Robotics of the Polish Academy of Sciences.

Ekemeyong Awong Lisiane Esther, MSc Eng.
ekemeyong.lisane_esther.dokt@pw.edu.pl
ORCID: 0000-0002-2047-0152

She is a PhD student at Warsaw University of Technology, Poland. She received her B.Tech degree in electronics and instrumentation engineering from National Institute of Technology Silchar, India and M.Eng degree in computer intelligent control and electromechanical engineering from Zhejiang Normal University (ZJNU), China. She co-authored one publication. Her research interest includes robotics and machine learning.