Intelligent Consumer Technologies

Prof. Paolo Napoletano
a.a. 2024/2025

Signal, image, and natural language processing in Consumer Technologies

Sensors in Consumer Technologies

Topics: Computer Vision, CV in Consumer Technologies, Face Detection, Face Recognition

Learning Objectives
- Definition of awareness in CT
- Definition of IMU
- Definition of HAR
- Definition of HAR pipeline

Sensors in Consumer Technologies

Some examples

Wearable (commercial and not) and ambient sensors of everyday life

Heart Rate - HR

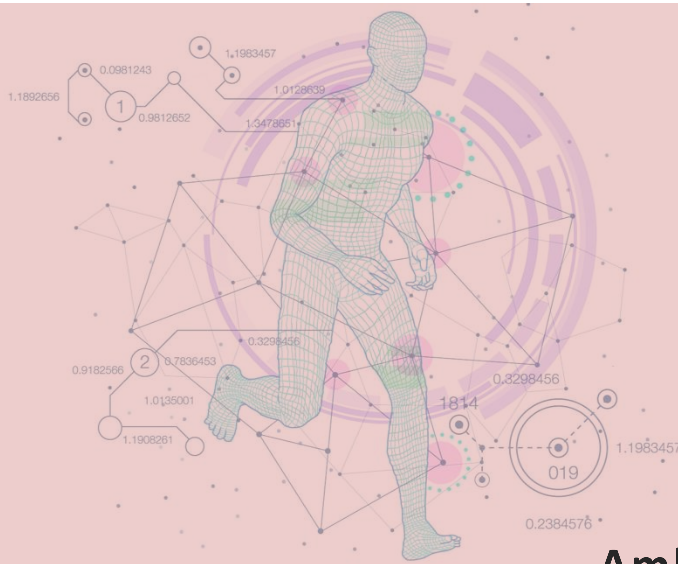
Electrodermal Activity - EDA

Electrocardiography - ECG

Electromyography - EMG

Heart Rate Variability - HRV

Electroencephalography - EEG



Breathing rate

Accelerometer

Gyroscope

Blood Pressure

Body temperature

Speech (audio)

Ambient video camera

Sensors in Consumer Technologies

Some examples

Wearable (commercial and not) and ambient sensors of everyday life

POSSIBLE APPLICATIONS

Heart Rate - HR

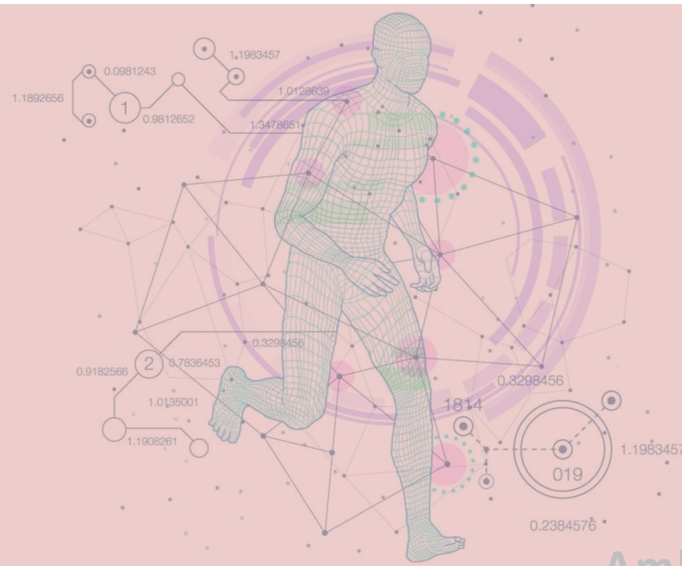
Electrodermal Activity - EDA

Electrocardiography - ECG

Electromyography - EMG

Heart Rate Variability - HRV

Electroencephalography - EEG



Breathing rate

Accelerometer

Gyroscope

Blood Pressure

Body temperature

Speech (audio)

Ambient video camera

HUMAN EMOTIONS: STRESS AND MENTAL WORKLOAD RECOGNITION

Sensors in Consumer Technologies

Some examples

Wearable (commercial and not) and ambient sensors of everyday life

POSSIBLE APPLICATIONS

Heart Rate - HR

Electrodermal Activity - EDA

Electrocardiography - ECG

Electromyography - EMG

Heart Rate Variability - HRV

Electroencephalography - EEG

Breathing rate

Accelerometer

Gyroscope

Blood Pressure

Body temperature

Speech (audio)

Ambient video camera

HUMAN BIOMETRIC: IDENTITY RECOGNITION AND VERIFICATION

Sensors in Consumer Technologies

Some examples

Wearable (commercial and not) and ambient sensors of everyday life

POSSIBLE APPLICATIONS

Heart Rate - HR

Electrodermal Activity - EDA

Electrocardiography - ECG

Electromyography - EMG

Heart Rate Variability - HRV

Electroencephalography - EEG

Breathing rate

Accelerometer

Gyroscope

Blood Pressure

Body temperature

Speech (audio)

Ambient video camera

HUMAN BIOMETRIC: IDENTITY RECOGNITION AND VERIFICATION

Sensors in Consumer Technologies

Some examples

Wearable (commercial and not) and ambient sensors of everyday life

POSSIBLE APPLICATIONS

Heart Rate - HR

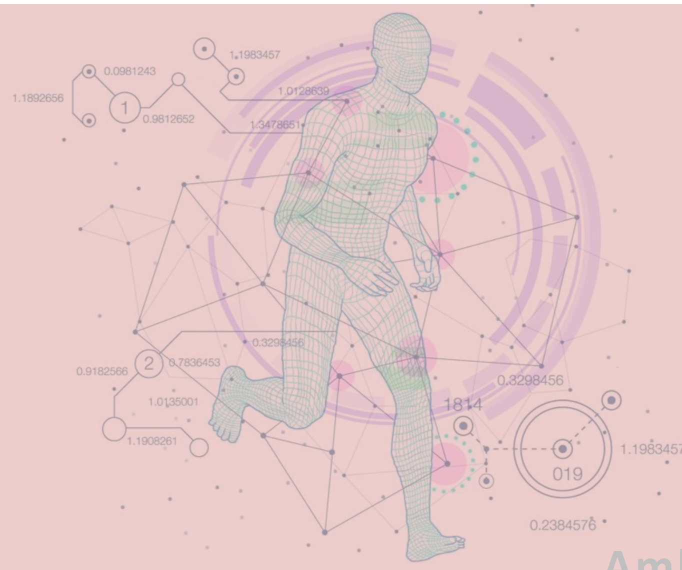
Electrodermal Activity - EDA

Electrocardiography - ECG

Electromyography - EMG

Heart Rate Variability - HRV

Electroencephalography - EEG



Breathing rate

Accelerometer

Gyroscope

Blood Pressure

Body temperature

Speech (audio)

Ambient video camera

HUMAN ACTION: FALL DETECTION AND ACTION RECOGNITION

Sensors in Consumer Technologies

Some examples

Wearable (commercial and not) and ambient sensors of everyday life

POSSIBLE APPLICATIONS

Heart Rate - HR

Electrodermal Activity - EDA

Electrocardiography - ECG

Electromyography - EMG

Heart Rate Variability - HRV

Electroencephalography - EEG

Breathing rate

Accelerometer

Gyroscope

Blood Pressure

Body temperature

Speech (audio)

Ambient video camera

HUMAN BIOMETRIC: IDENTITY RECOGNITION AND VERIFICATION

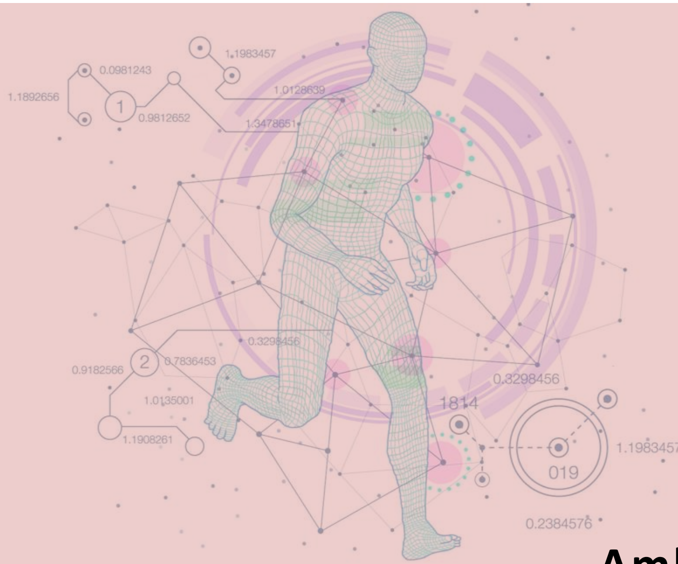
Sensors in Consumer Technologies

Some examples

Wearable (commercial and not) and ambient sensors of everyday life

POSSIBLE APPLICATIONS

Heart Rate - HR
Electrodermal Activity - EDA
Electrocardiography - ECG
Electromyography - EMG
Heart Rate Variability - HRV
Electroencephalography - EEG



Breathing rate
Accelerometer
Gyroscope
Blood Pressure
Body temperature
Speech (audio)

Ambient video camera

HUMAN ACTION: FALL DETECTION AND ACTION RECOGNITION

Wearable (commercial and not) and ambient sensors of everyday life

POSSIBLE APPLICATIONS

Heart Rate - HR

Electrodermal Activity - EDA

Electrocardiography - ECG

Electromyography - EMG

Heart Rate Variability - HRV

Electroencephalography - EEG

Breathing rate

Accelerometer

Gyroscope

Blood Pressure

Body temperature

Speech (audio)

Ambient video camera

HUMAN EMOTIONS: SHORT-TERM AND LONG-TERM MOOD

Current consumer electronics are more **connected** than **smart**

Internet of Things (IoT)

We want to take a **step forward**

We want to put **Artificial Intelligence** (**AI**) inside **CE**

(IoT)

… and make them **connected** and **aware of us**

Sensors in Consumer Technologies

Awareness

Hello Google

Hello

Can you turn off the light in the living room?

Yes, I can

Sensors in Consumer Technologies

Awareness

Hello Google

Hello

Can you turn off the light in the living room?

Yes, I can

BEING MORE AWARE OF THE USER

Sorry, I can’t. You are unauthorized for this task. Please ask Paolo, the owner of this device, to add you to the list of authorized people

Yes Paolo, I do it.
I see you quite tired, do you want me to put some relaxing music on?

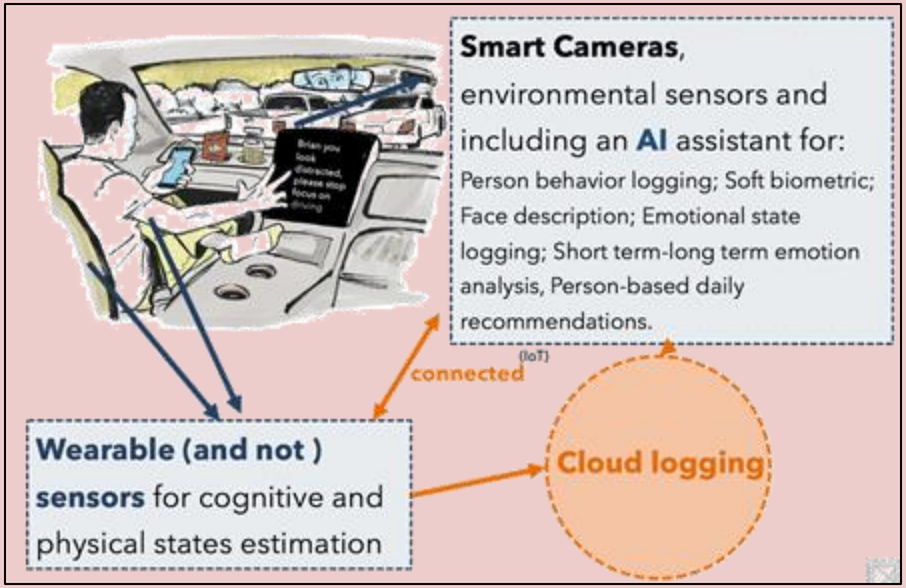
Scenarios Consumer Technologies

Examples

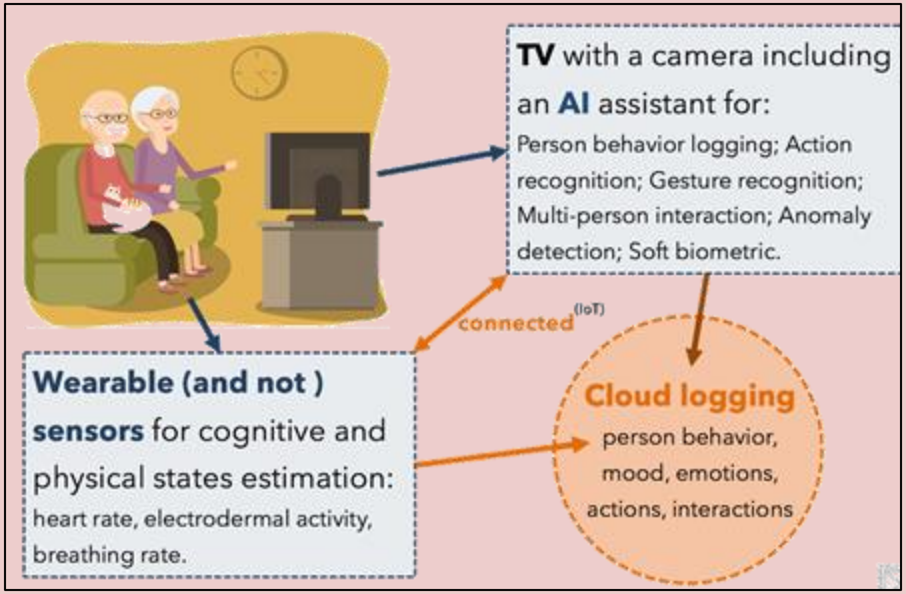
The ultimate scope is to make consumer electronics more aware of their users in terms of **physical** characteristics, **emotions**, **mood** and **habits**

Possible scenarios

Vehicle Driver Monitoring (ADAS Level 0-4)



Assisted Living Room



Enabling technologies: **activity monitoring, identity recognition, emotion recognition, mood monitoring, soft biometric**

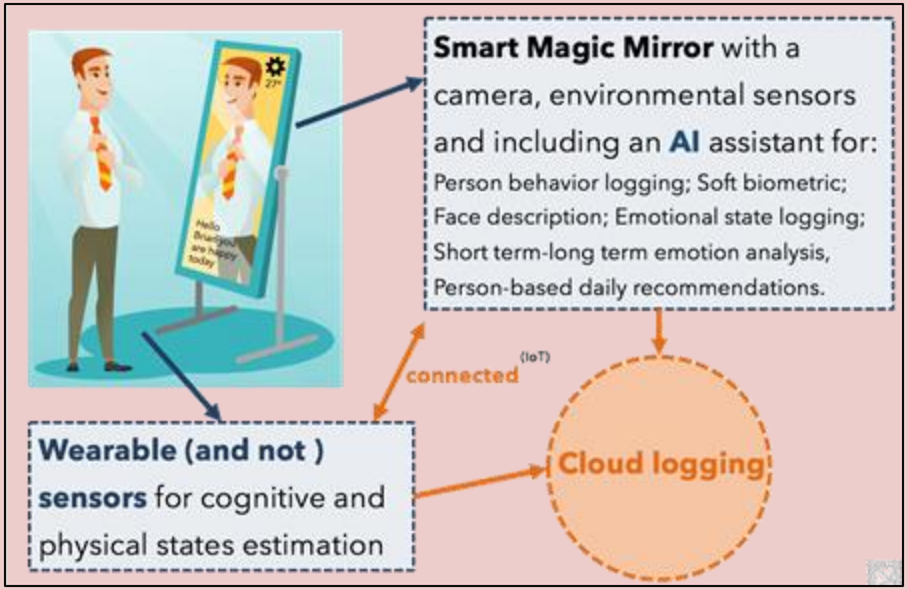
Scenarios Consumer Technologies

Examples

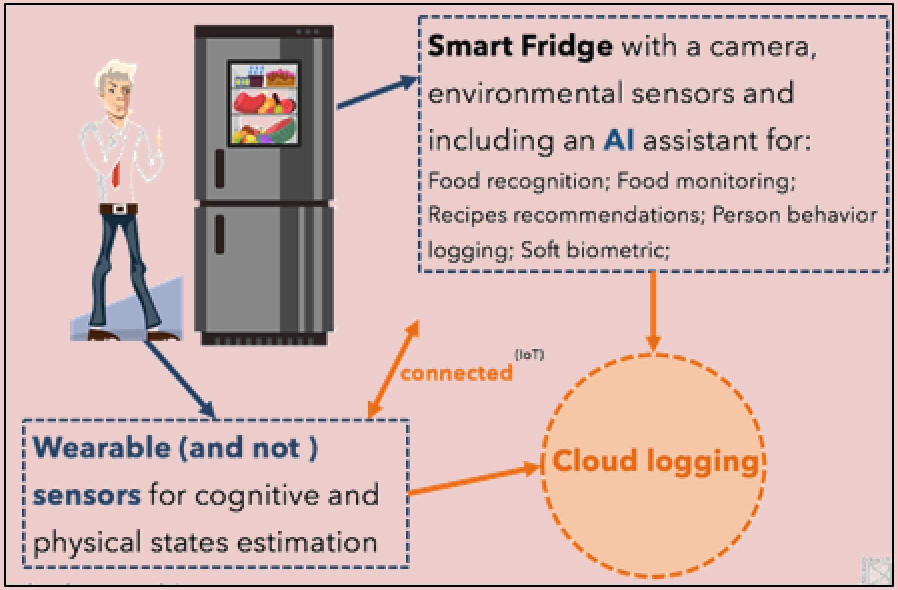
The ultimate scope is to make consumer electronics more aware of their users in terms of **physical** characteristics, **emotions**, **mood** and **habits**

Possible scenarios

Smart magic mirror



Smart Kitchen

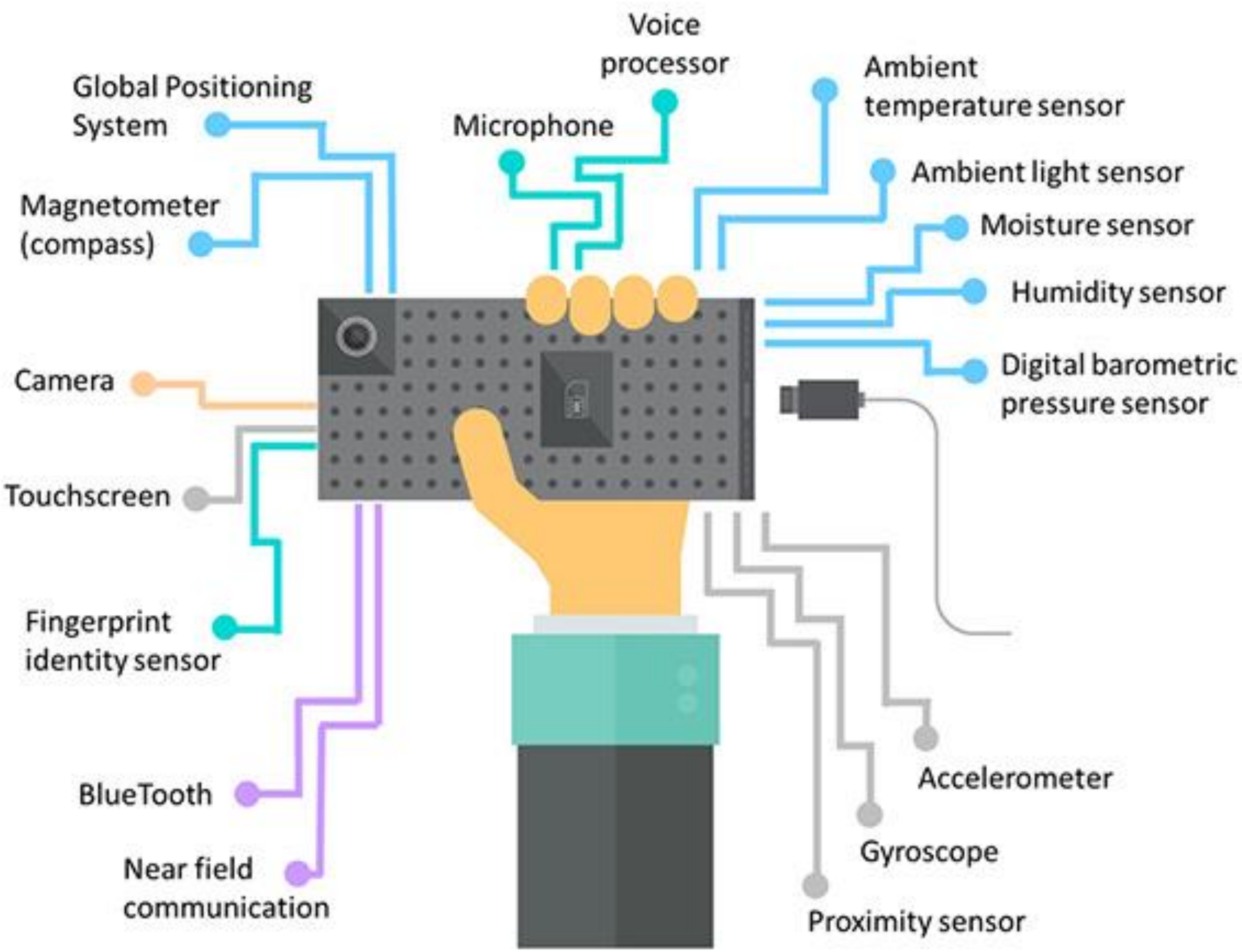


Enabling technologies: **activity monitoring, identity recognition, emotion recognition, mood monitoring, soft biometric**

Sensors in Consumer Technology

Sensors in Consumer Technologies

Some examples



Sensors in Consumer Technologies

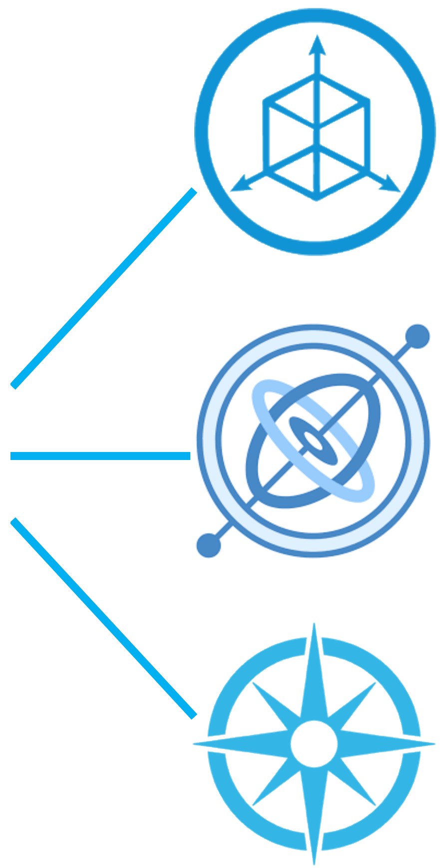
Some examples

- **iPhone's sensors (14)**
  - Face ID
  - Barometer
  - High dynamic range gyro
  - High-g accelerometer
  - Proximity sensor
  - Dual ambient light sensors

- **Apple Watch's sensors (9):**
  - Third-generation optical heart sensor
  - Temperature sensing
  - Compass with Waypoints and Backtrack
  - Always-on altimeter
  - High-g accelerometer
  - High dynamic range gyroscope
  - Ambient light sensor

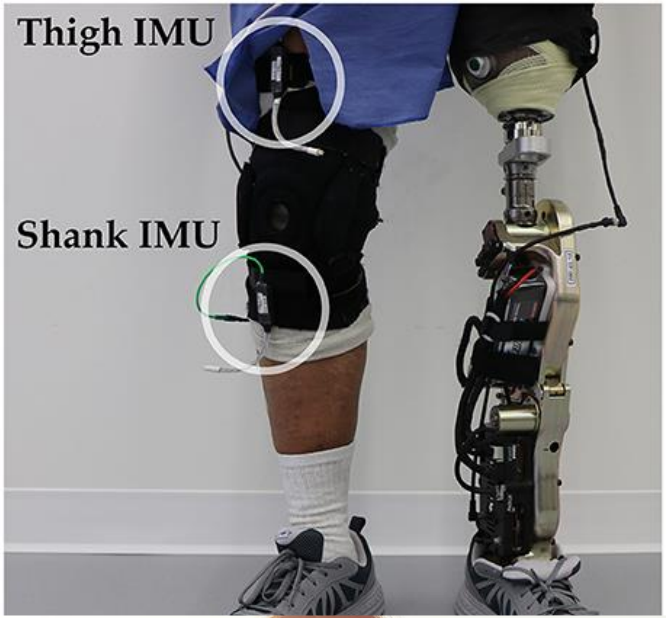
Inertial Measurement Unit (IMU)

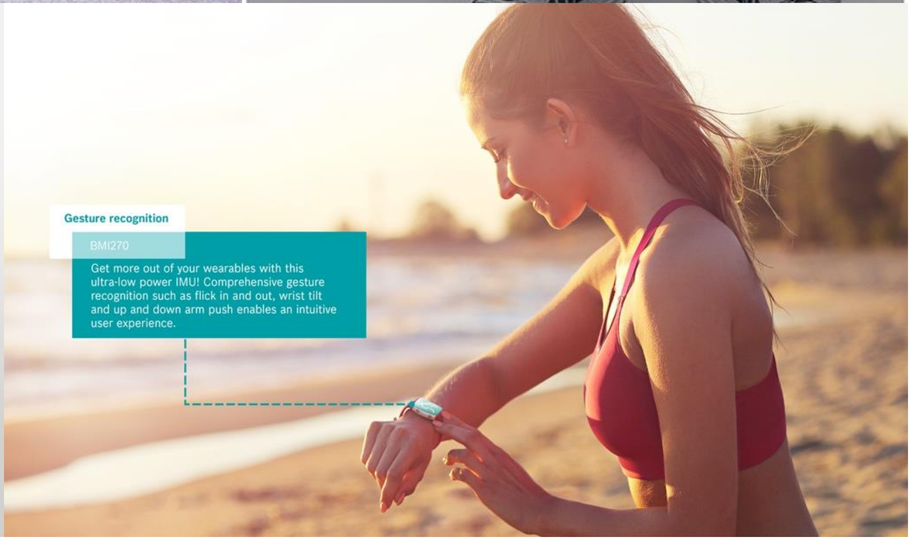
Definition



- Accelerometer
- Gyroscope
- Magnetometer (compass)

IMU: Applications



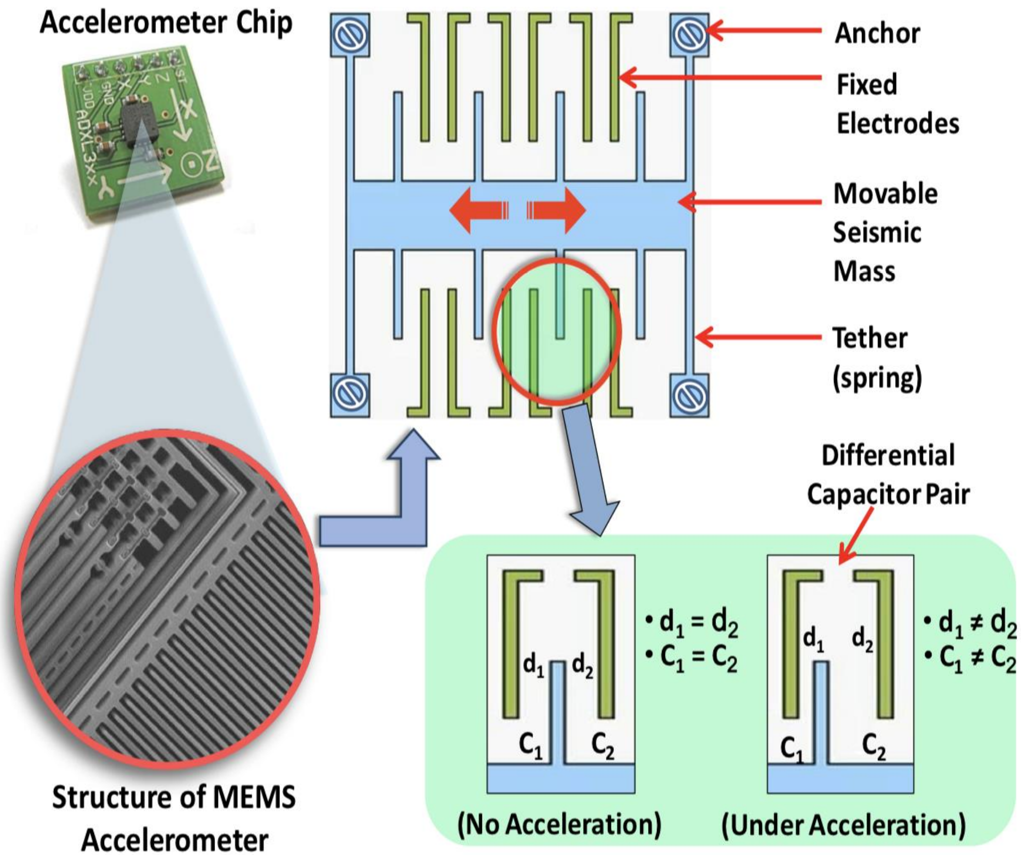


Accelerometer

Explanation

Measures acceleration, not speed.

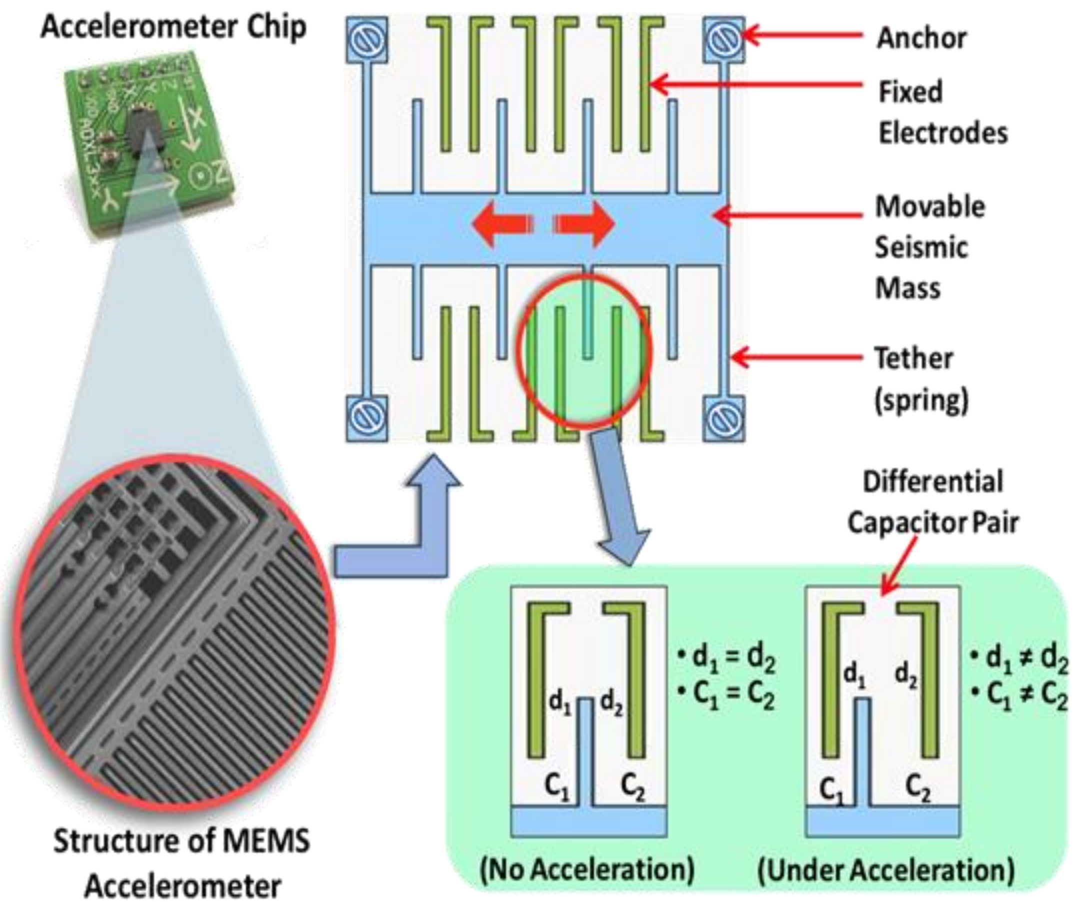
Measurements are proportional to the force. Why?



MEMS - micro- electro-mechanical systems

An **accelerometer** is an electromechanical sensor that captures the rate of change of the velocity of an object over time.

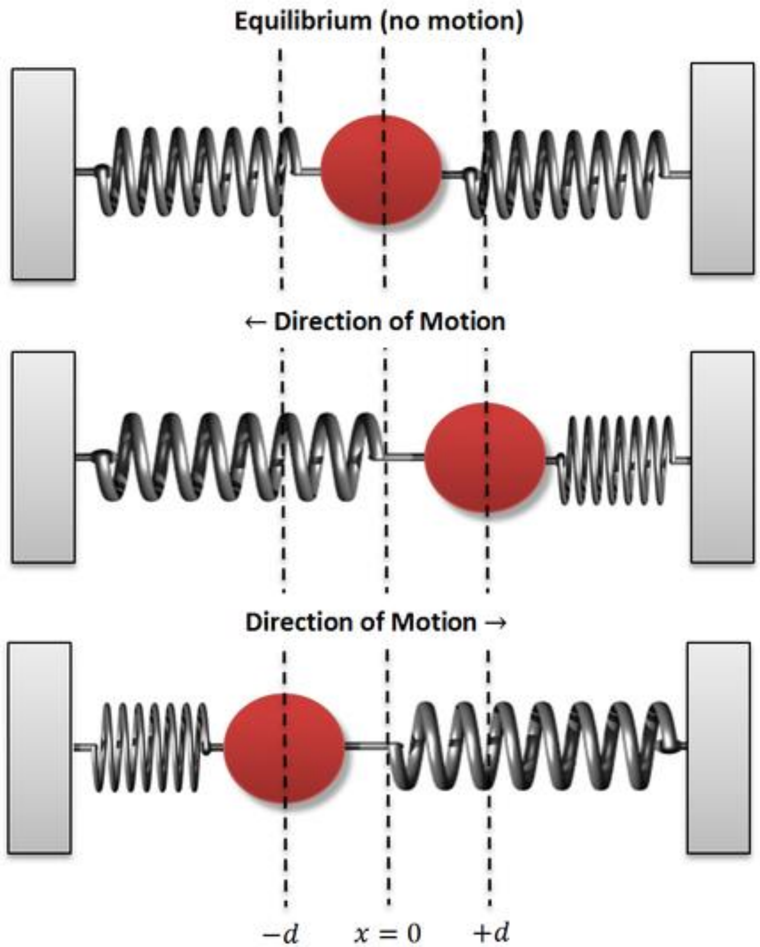
It records **three-dimension acceleration**, which joins the reference devices' axes. Thus, a trivariate time series is produced. The measuring unit is m/s^2 or g forces.



MEMS - micro- electro-mechanical systems

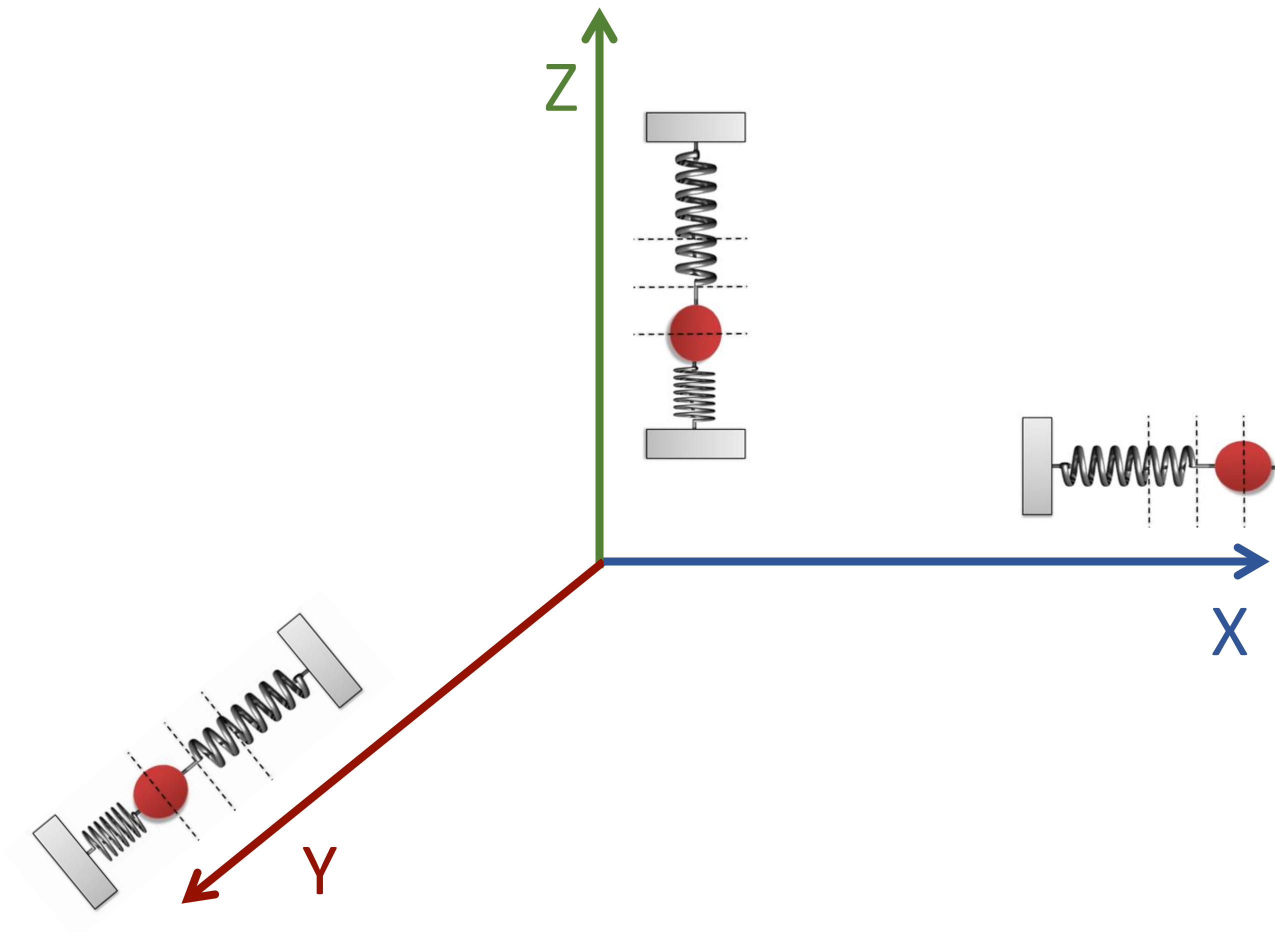
Accelerometer

Explanation



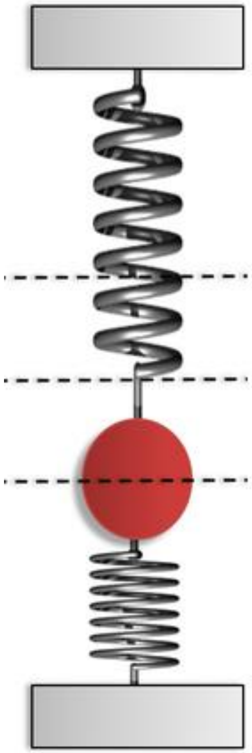
Accelerometer

Explanation



Accelerometer

Explanation



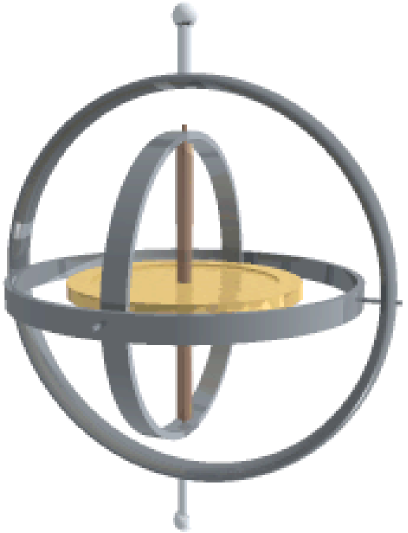
Gravity = 9.8 m/s^2

What reading will you get when the device is in “free-fall”?

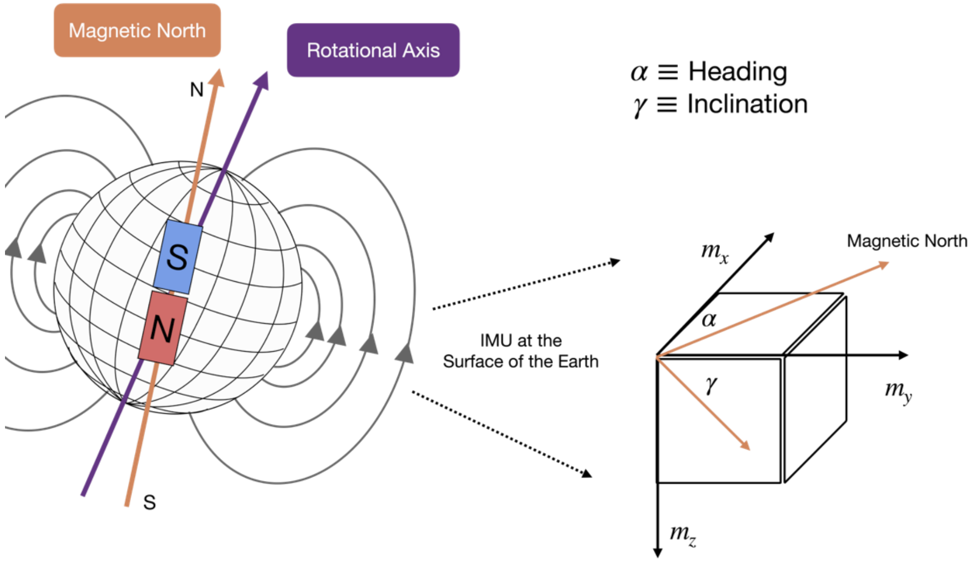
IMU

Explanation

The **gyroscope** measures three-axial angular velocity. Its unit is measured in degrees over seconds (*degrees/s*).



A **magnetometer** measures the change of a magnetic field at a particular location. The measurement unit is *Tesla* (T) and is usually recorded on the three axes.

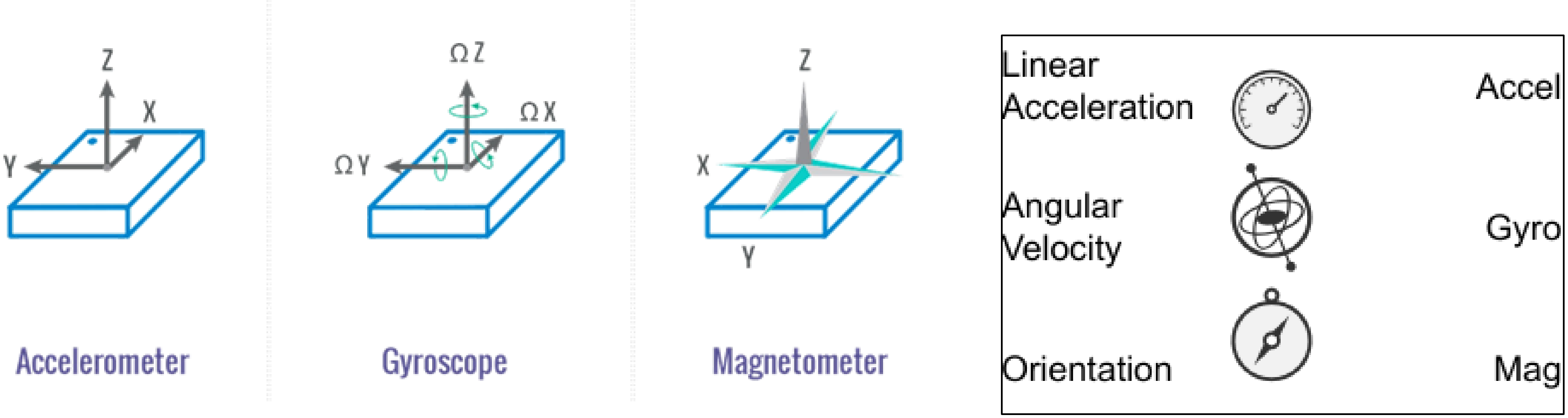


Roll, Pitch and Yaw

Explanation

Pitch, **roll** and **yaw** angles are one of the most important pieces of information extracted from the IMU sensor output.

- **Pitch** angle is formed between the nose of the body and the flat surface.
- **Roll** angle is formed between the body and the flat surface by rolling the body left and right.
- **Yaw** angle refers to the rotation with respect to the vertical axis

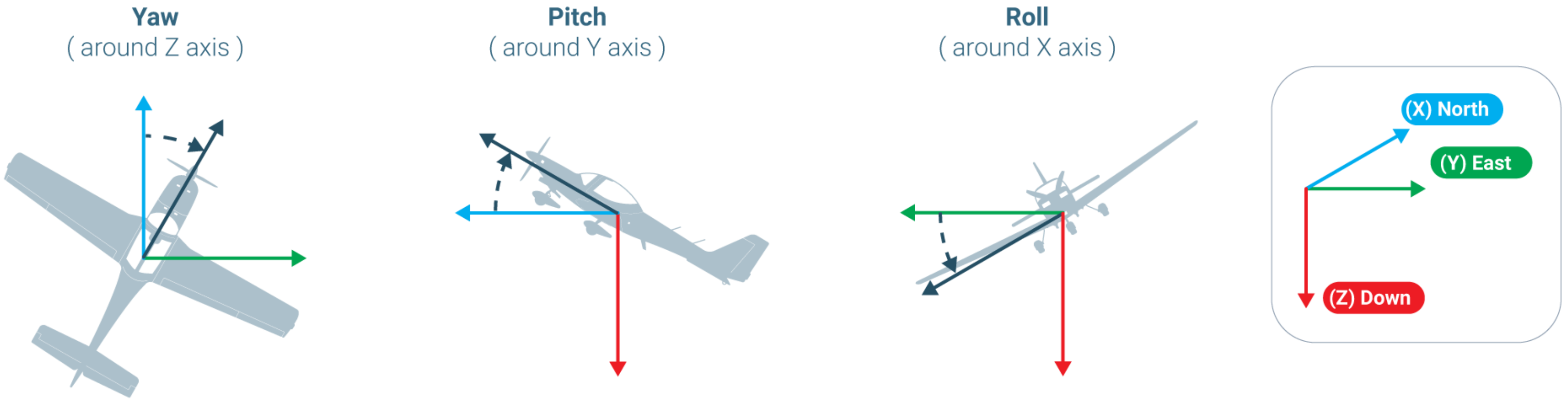


Roll, Pitch and Yaw

Explanation

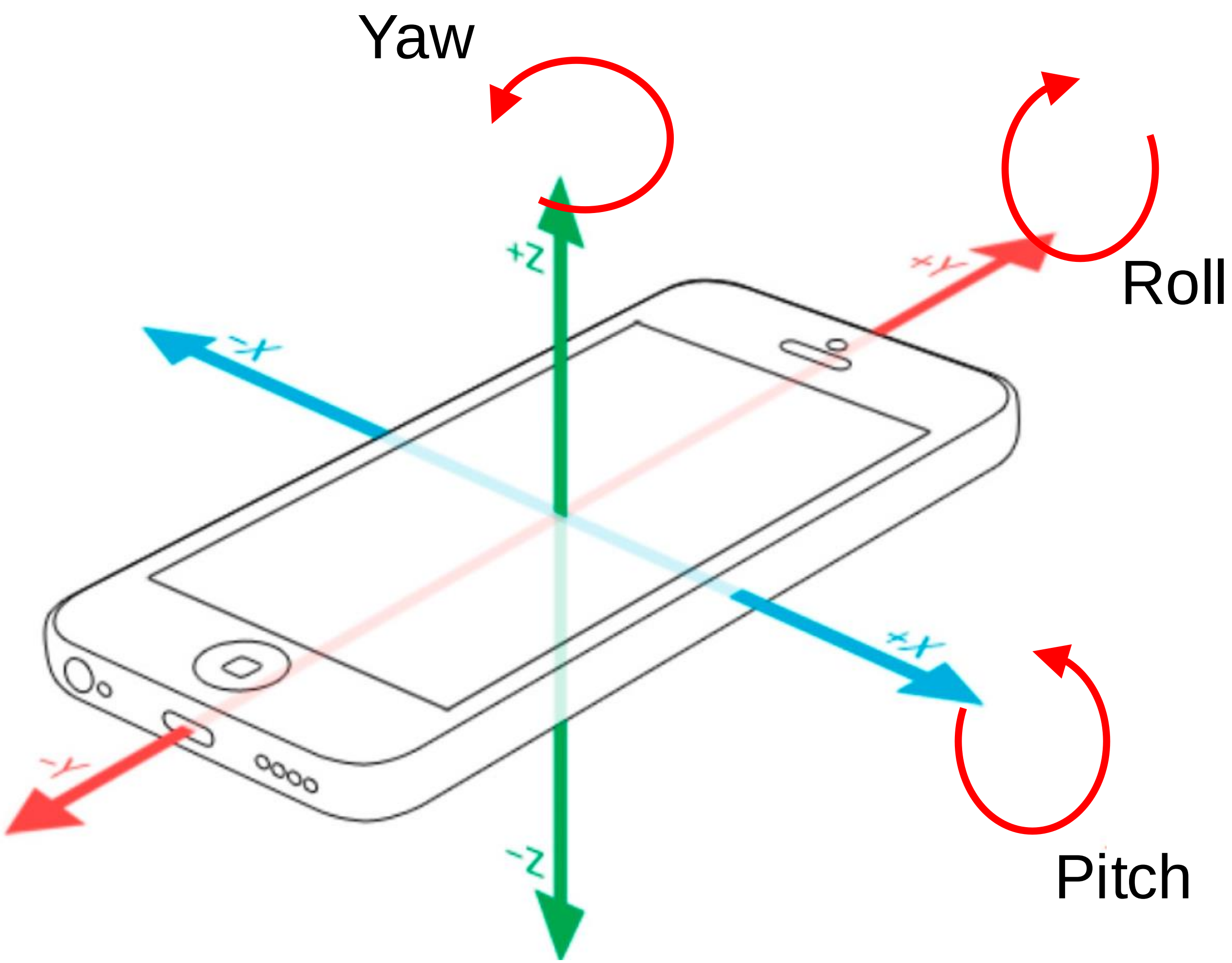
Pitch, **roll** and **yaw** angles are one of the most important pieces of information extracted from the IMU sensor output.

- **Pitch** angle is formed between the nose of the body and the flat surface.
- **Roll** angle is formed between the body and the flat surface by rolling the body left and right.
- **Yaw** angle refers to the rotation with respect to the vertical axis



Pitch-Roll-Yaw on a Smartphone

Explanation

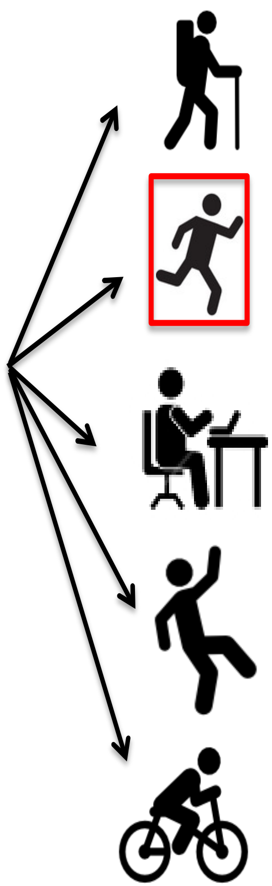


Human Activity Recognition

Human Activity Recognition

Definition

- Recognizing **human activities** and monitoring population behavior are fundamental needs of our society.
- Population **security**, **crowd surveillance**, **healthcare** support and **living assistance**, and lifestyle and behavior tracking are some of the main applications that require the recognition of human activities.
- **Human Activity Recognition** (HAR) is the field of research which focuses on all these tasks. **HAR** automatically recognizes human activity by analyzing signals acquired by sensors: : **environmental** and **wearable**.



* See **additional materials** on Ferrari, A., Micucci, D., Mobilio, M., & Napoletano, P. (2021). Trends in human activity recognition using smartphones. *Journal of Reliable Intelligent Environments*, *7*(3), 189-213.

Human Activity Recognition

Definition

- Among the environmental devices, cameras are the most used, while wearable devices encompass all on body worn sensors, such as smart-shirt, smart-shoes, ad-hoc **Inertial Measurement Unit** (IMU), smartphones and smartwatches.

- Nowadays, commercial fitness **bracelets**, **smartwatches**, and **smartphones** are equipped with inertial sensors, such as **accelerometer** and **gyroscope**.

- Such sensors acquire signals that are exploited by machine learning methods for automatic **Human Activity Recognition** (HAR)

* See **additional materials** on Ferrari, A., Micucci, D., Mobilio, M., & Napoletano, P. (2021). Trends in human activity recognition using smartphones. *Journal of Reliable Intelligent Environments*, *7*(3), 189-213.

Human Activity Recognition

Definition

- Human Activity Recognition (**HAR**): methods and techniques to automatically identify Activities of Daily Living (**ADLs**) from inertial sensors of smartphones and smartwatches

- Recent statistics show that more than 5.19 billion people (67% of the world's population) use **smartwatches**, with user numbers up by 124 million (2.4%) over the past year.

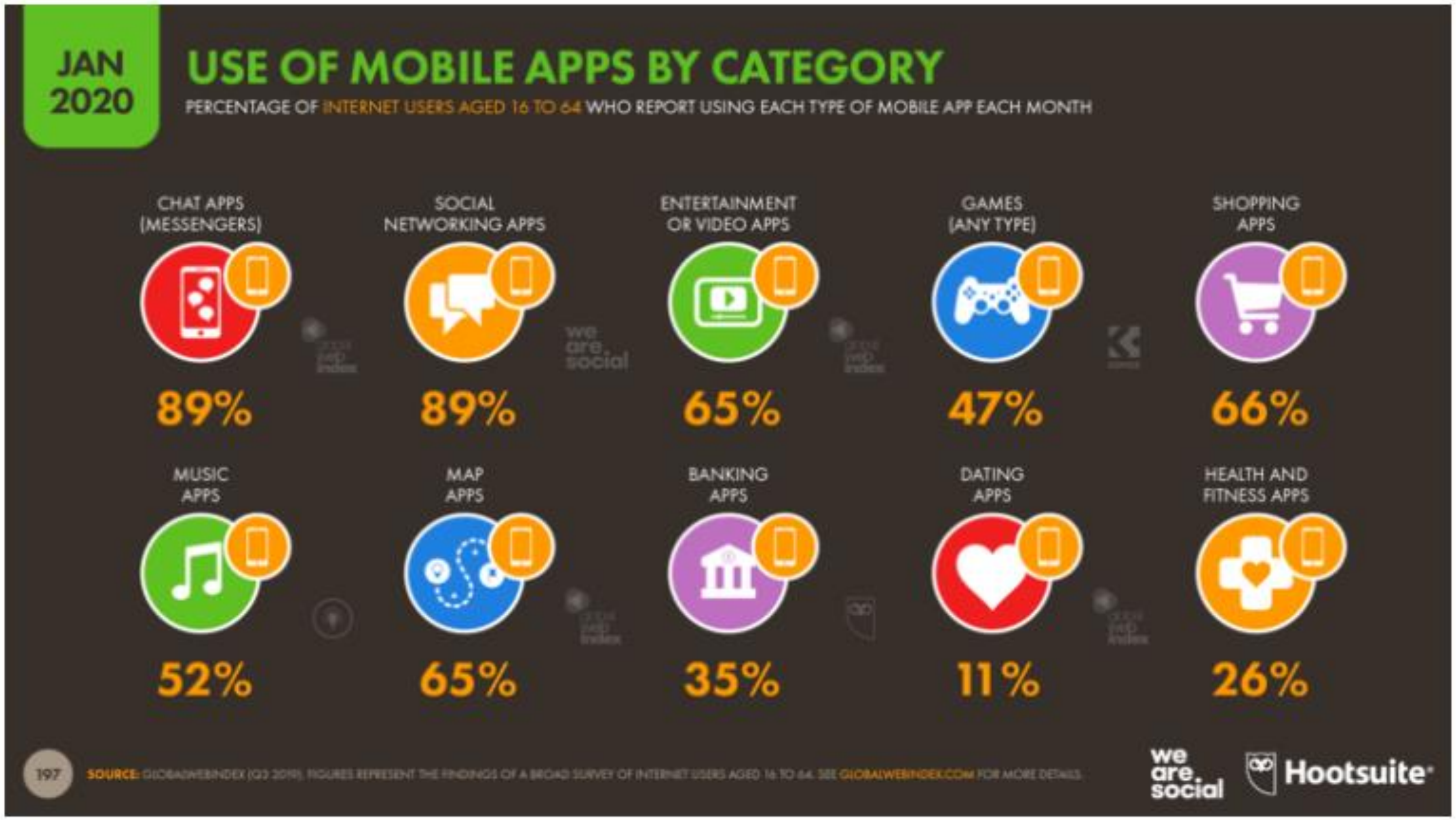
- HAR can be extensively used in several domains. For example, in **healthcare** to keep track of elderly people or the rehabilitation process after accidents and injuries, or in the guide to safe travel

* See **additional materials** on https://datareportal.com/reports/digital-2020-global-digital-overview

Use of mobile phone

Numbers

People use apps in almost every aspect of their lives, whether it's staying in touch with friends and family, relaxing on the couch, managing their finances, getting fit, or even finding love.



* See **additional materials** on Use of Mobile Apps by Category January 2020. Source: DataReportal

Activity Monitoring

Types of actions

Activity monitoring

Distress detection

Anomaly detection

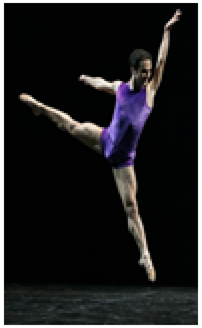
HAR

Challenges

Atomic / complex actions

JUMP

VS



DANCE

Single subject / multi-subject

RUNNING

VS

CHASING

Interaction with objects

DRINKING

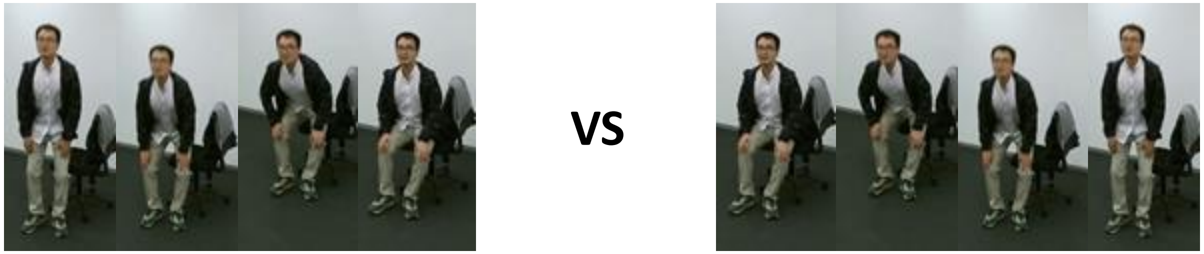
VS

SINGING

HAR

Challenges

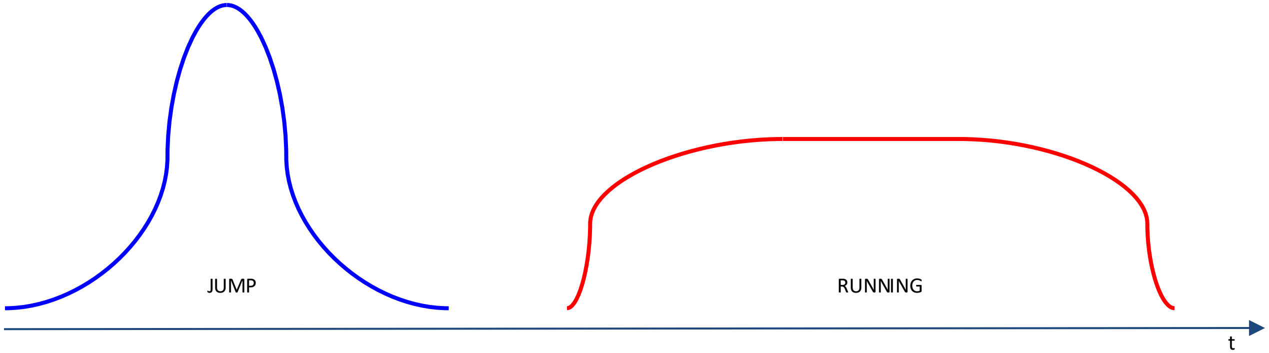
Order matters



SITTING DOWN

STANDING UP

Temporal saliency



HAR – Activity of Daily Living

Type of actions

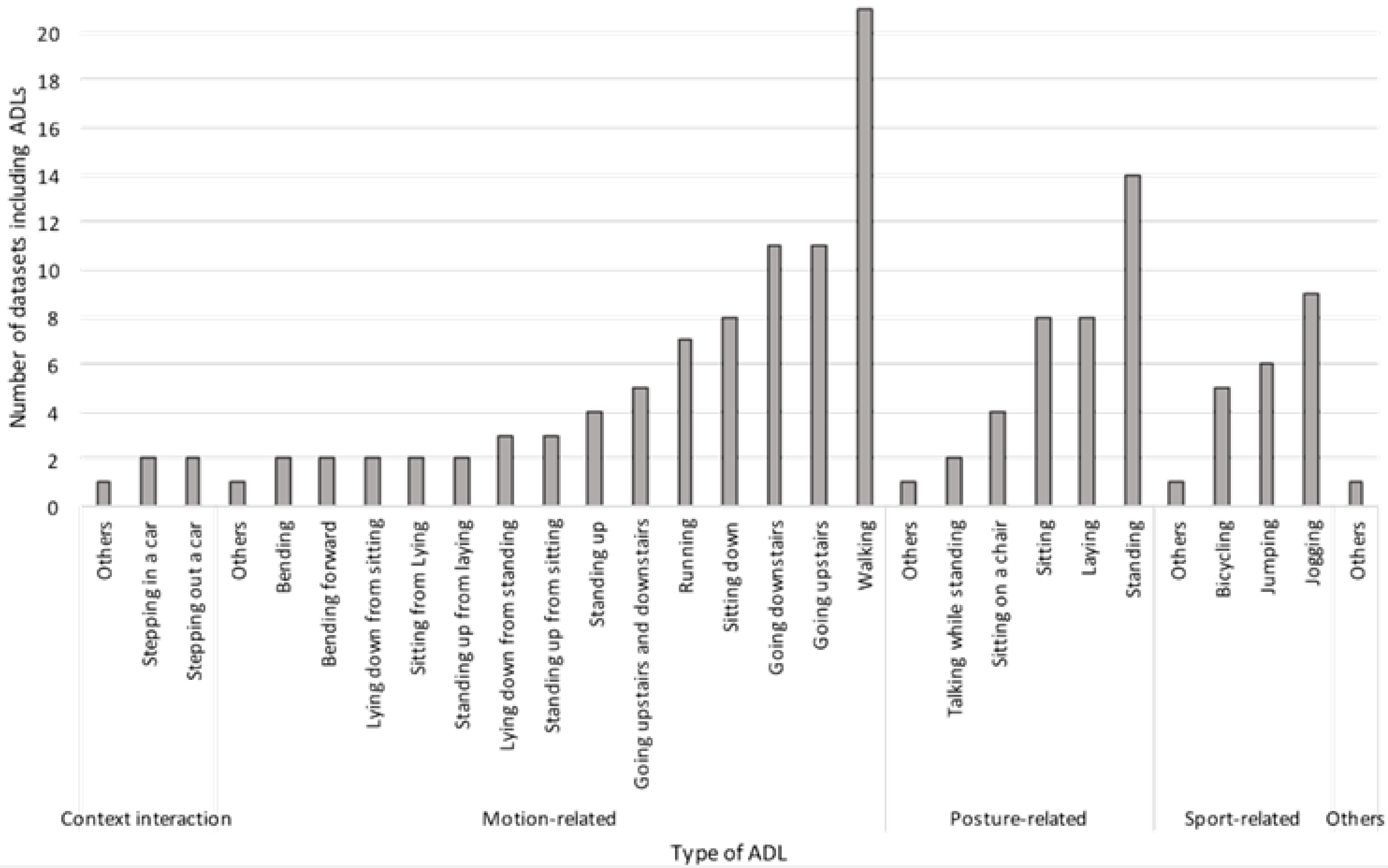
The activities of daily living (**ADLs**) is a term used to collectively describe fundamental skills that are required to independently care for oneself such as eating, bathing, and mobility. The term activities of daily living was first coined by Sidney Katz in 1950.

* See **additional materials** on Edemekong, P. F., Bomgaars, D., Sukumaran, S., & Levy, S. B. (2019). Activities of daily living.

HAR – Activity of Daily Living

Type of actions

Recognition of activities of daily living (**ADLs**) may allow to infer the amount of physical activity that a subject perform daily.

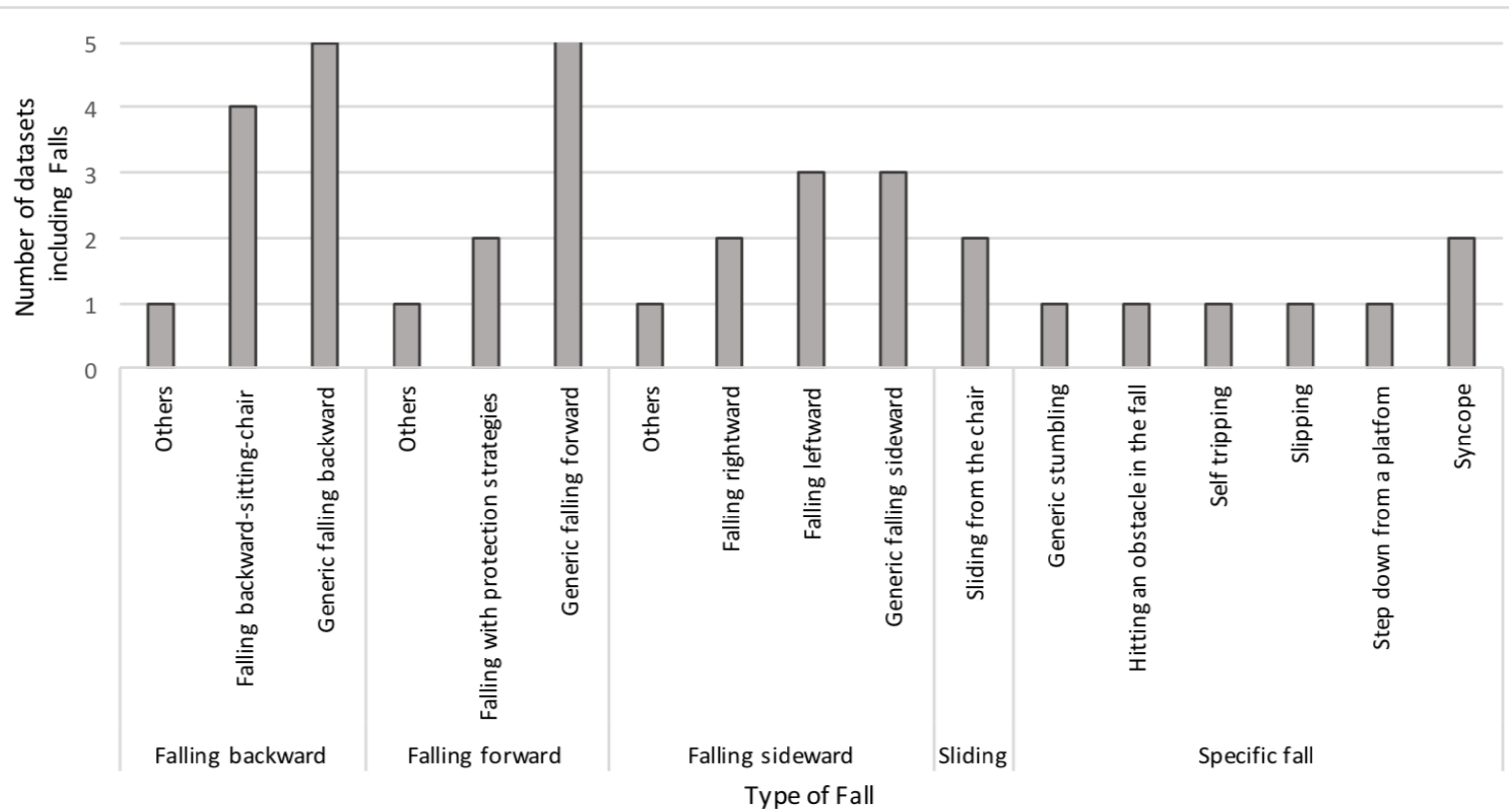


* See **additional materials** on Micucci, D., Mobilio, M., & Napoletano, P. (2017). Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences*, *7*(10), 1101.

HAR – Activity of Daily Living

Type of actions

Falls and theirs occurrence in the publicly available datasets analysed grouped by category.



* See **additional materials** on Micucci, D., Mobilio, M., & Napoletano, P. (2017). Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences*, *7*(10), 1101.

HAR – Activity of Daily Living

DBs

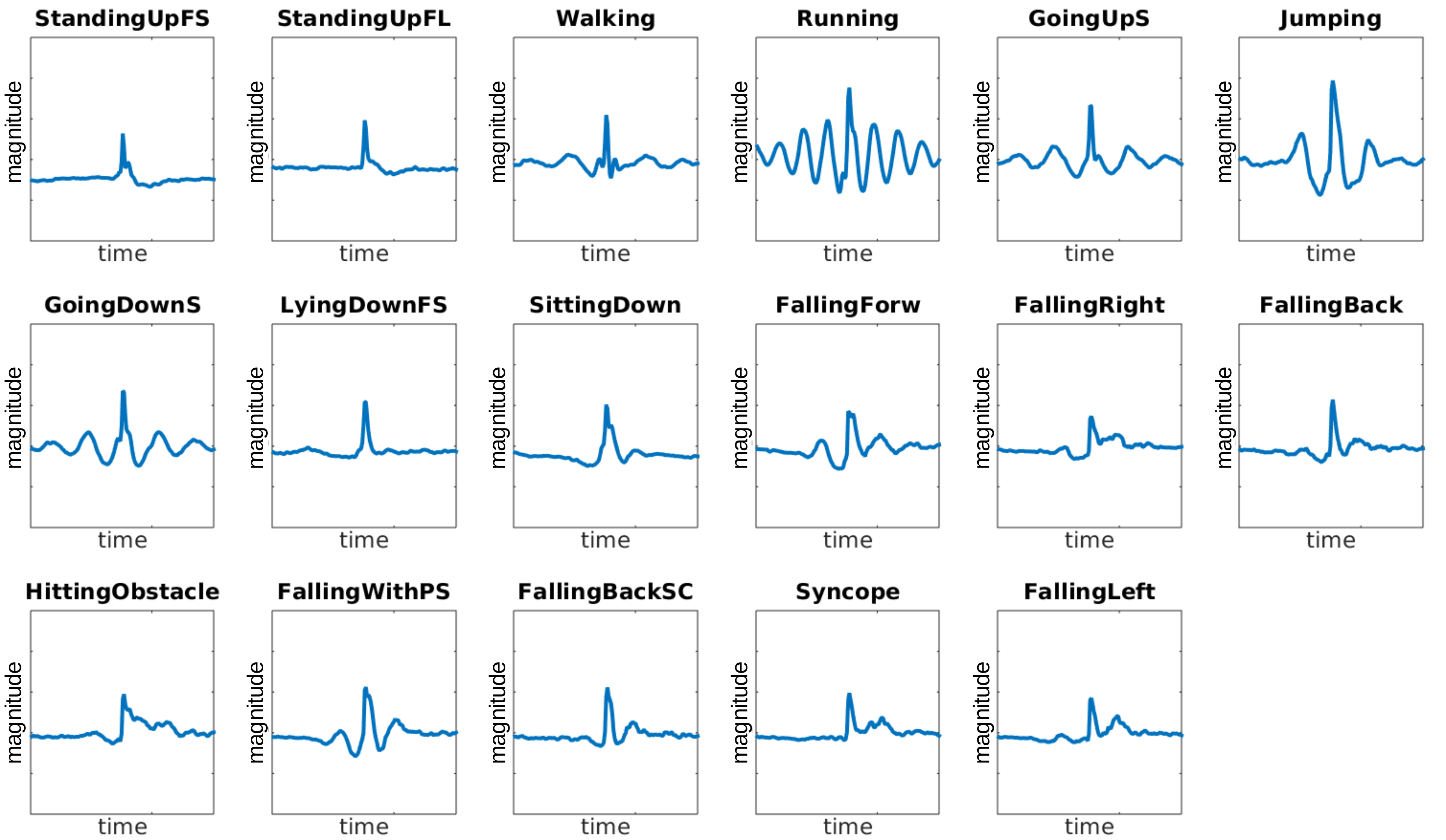
Publicly available datasets containing samples from smartphones sensors.

Dataset	Year	ADLs	Falls	Nr. of Subjects	Gender Female	Gender Male	Age (Years)	Height (cm)	Weight (Kg)
DMPSBFD [24]	2015	yes	yes	5	-	-	-	-	-
Gravity [23]	2016	yes	yes	2	-	-	26–32 29 ± 4.2	170–185 178 ± 10.6	63–80 71.5 ± 12
MobiFall [11]	2014	yes	yes	24	7	17	22–47 27 ± 5	160–189 175 ± 7	50–103 76.4 ± 14.5
MobiAct [25]	2016	yes	yes	57	15	42	20–47 25 ± 4	160–193 175 ± 4	50–120 76.6 ± 14.4
RealWorld (HAR) [26]	2016	yes	no	16	7	8	16–62 32 ± 12	163–183 173 ± 7	48–95 74.1 ± 13.3
Shoaib PA [27]	2013	yes	no	4	0	4	25–30 -	-	-
Shoaib SA [28]	2014	yes	no	10	0	10	25–30 -	-	-
tFall [29]	2013	yes	yes	10	7	3	20–42 31 ± 9	161–184 173 ± 1	54–98 69.2 ± 13.1
UCI HAR [30]	2012	yes	no	30	-	-	19–48 -	-	-
UCI HAPT [31]	2015	yes	no	30	-	-	19–48 -	-	-
UCI UIWADS [32]	2013	yes	no	22	-	-	- -	-	-
UMA Fall [33]	2016	yes	yes	17	6	11	14–55 27 ± 10	155–195 172 ± 9	50–93 69.9 ± 12.3
WISDM [34]	2012	yes	no	29	-	-	-	-	-
UniMiB SHAR	2016	yes	yes	30	24	6	18–60 27 ± 11	160–190 169 ± 7	50–82 64.4 ± 9.7

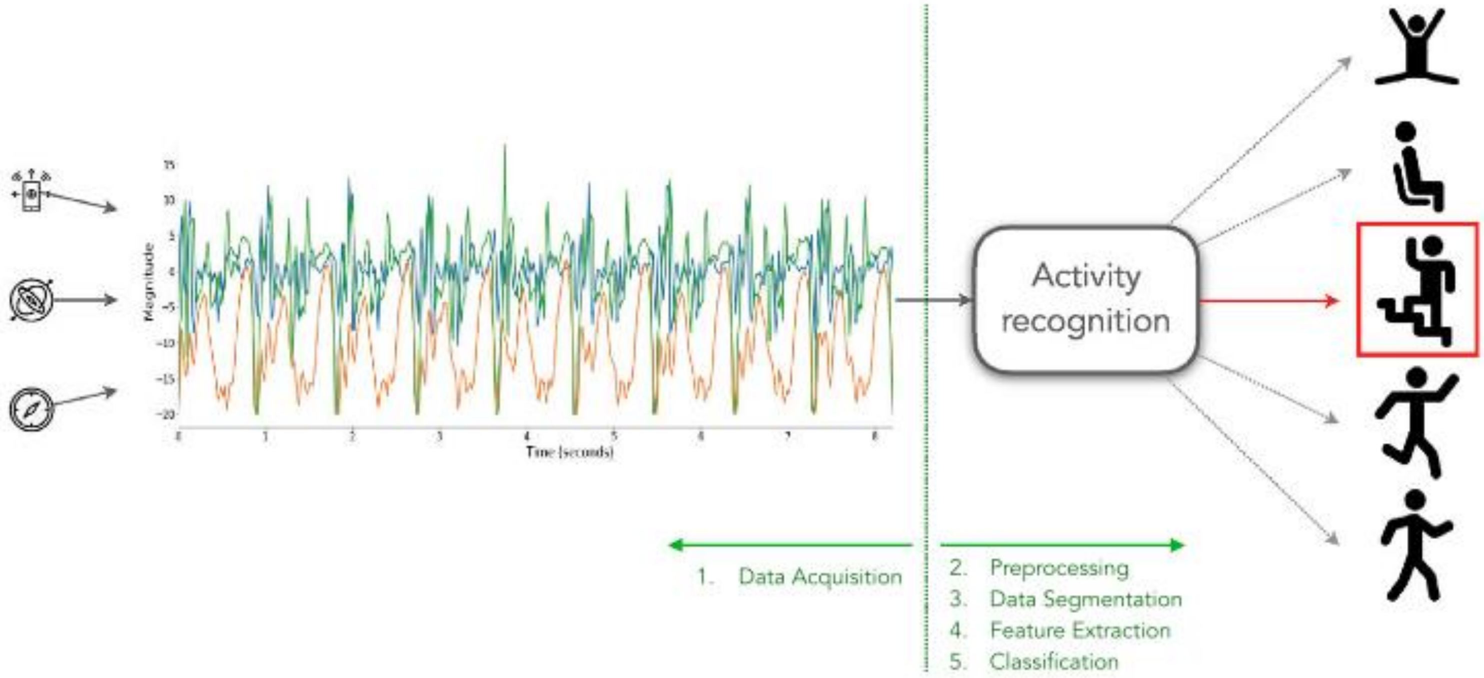
* See **additional materials** on Micucci, D., Mobilio, M., & Napoletano, P. (2017). Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences*, *7*(10), 1101.

HAR – Samples

Accelerometer



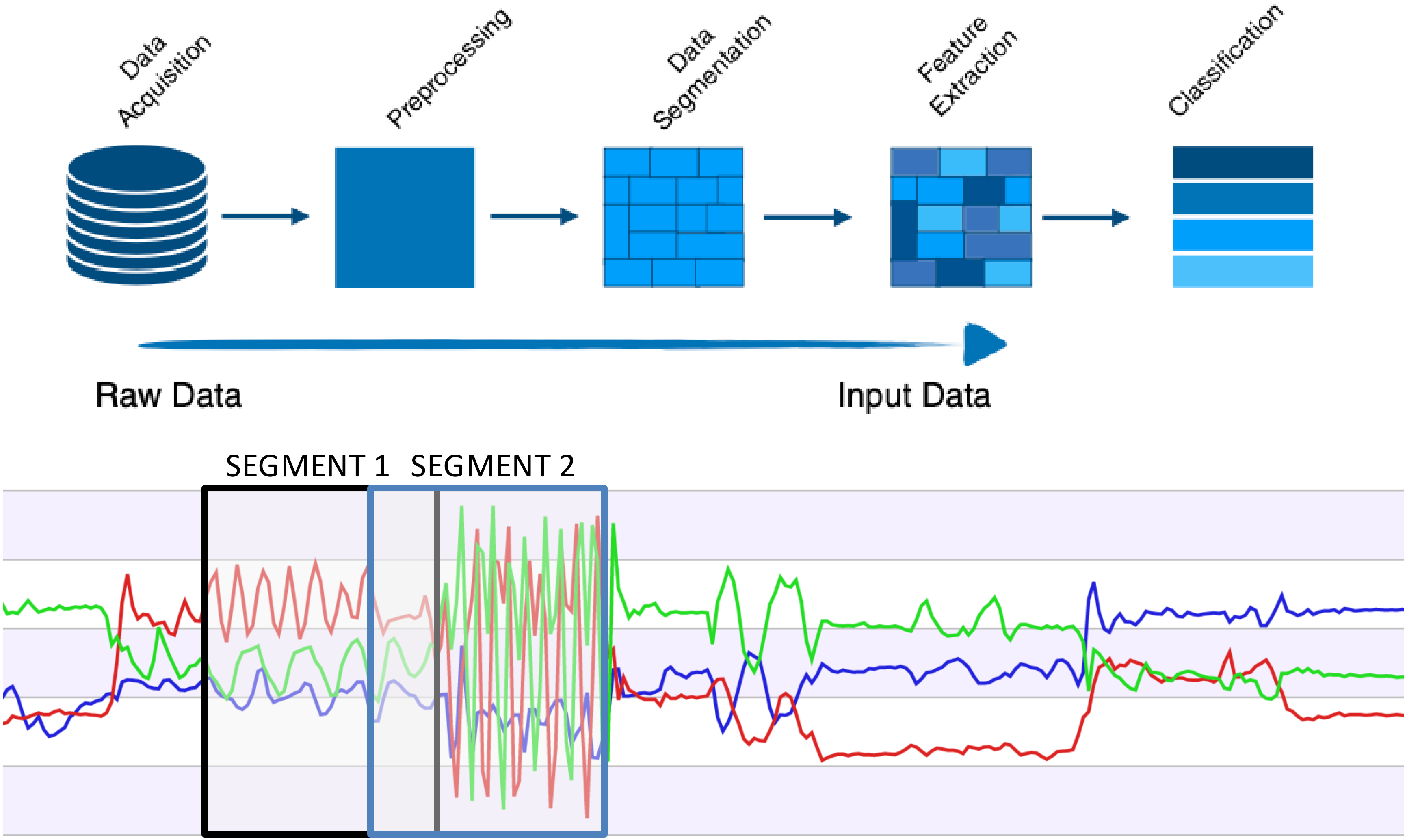
An abstracted overview of the human activity recognition process



* See **additional materials** on Ferrari, A., Micucci, D., Mobilio, M., & Napoletano, P. (2021). Trends in human activity recognition using smartphones. *Journal of Reliable Intelligent Environments*, *7*(3), 189-213.

HAR

pipeline



* See **additional materials** on Ferrari, A., Micucci, D., Mobilio, M., & Napoletano, P. (2021). Trends in human activity recognition using smartphones. *Journal of Reliable Intelligent Environments*, *7*(3), 189-213.

HAR – Data Acquisition

Accelerometer and Gyroscope

- **Accelerometer** is the most popular sensor in HAR, because it measures the directional movement of a subject's motion status over time
- Nevertheless, it **struggles** to resolve lateral **orientation** or **tilt**, and to find out the location of the user, which are precious information for activity recognition.
- For these reasons, some sensor combinations have been proposed as valid solution in HAR. In most of the cases, **accelerometer** and **gyroscope** are used conjointly to both acquire more information about the device movements, and to possibility infer the device position

* See **additional materials** on https://datareportal.com/reports/digital-2020-global-digital-overview

HAR – Data Acquisition

Sampling rate

An important factor to consider in the acquisition step is the **sampling rate** that influences the number of available samples for the classification step.

The sampling rate is defined as the **number of data points** recorded **in a second** and is expressed in Hertz. For instance, if the sampling rate is equal to 50Hz, it means that 50 values per second are recorded.

In the literature, different sampling rates have been considered ranging from **30Hz** to **100Hz**

BOSCH	BNO055 **Data sheet**	Page 31

3.6.3 Fusion Output data rates

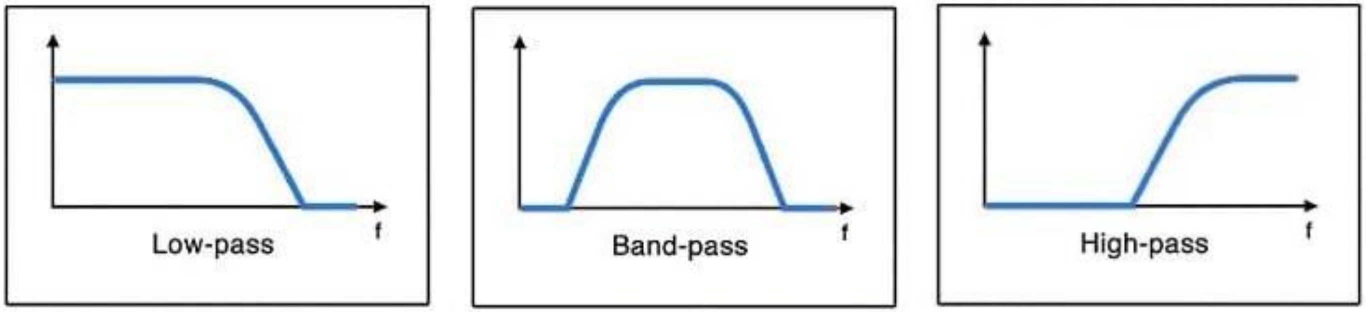
Table 3-14: Fusion output data rates

BNO055 Operating Mode	Data input rate			Algo calling rate	Data output rate			
	Accel	Mag	Gyro		Accel	Mag	Gyro	Fusion data
IMU	100Hz	NA	100Hz	100Hz	100Hz	NA	100Hz	100Hz

HAR – Data preprocessing

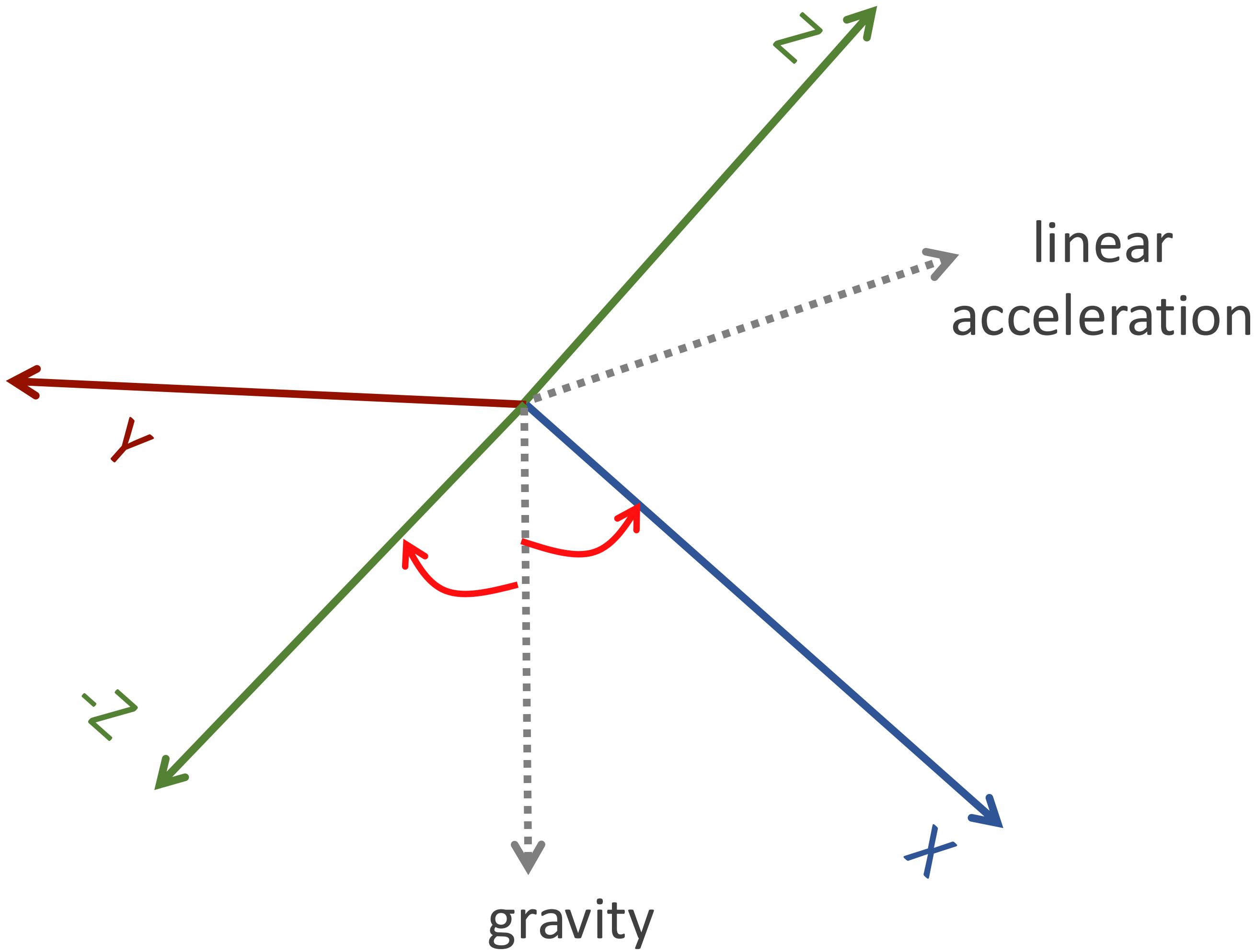
Filtering

- The **high-frequency** component of the accelerometer signal is mostly related to the action performed by the subjects, while the **low-frequency** component of the accelerometer signal is mainly related to the presence of **gravity**.
- Usually, a low-pass filter with **cut-off frequency** ranging between 0.1 and 0.5 Hz is used to isolate the gravity component. To find the body acceleration component, the result of the low-pass filtered signal is subtracted from the original signal.
- Filtering is also used to **clear raw data from artifacts**. It is stated that a cut-off frequency of 15Hz is enough to capture human body motion whose energy spectrum lies between 0 Hz and 15 Hz



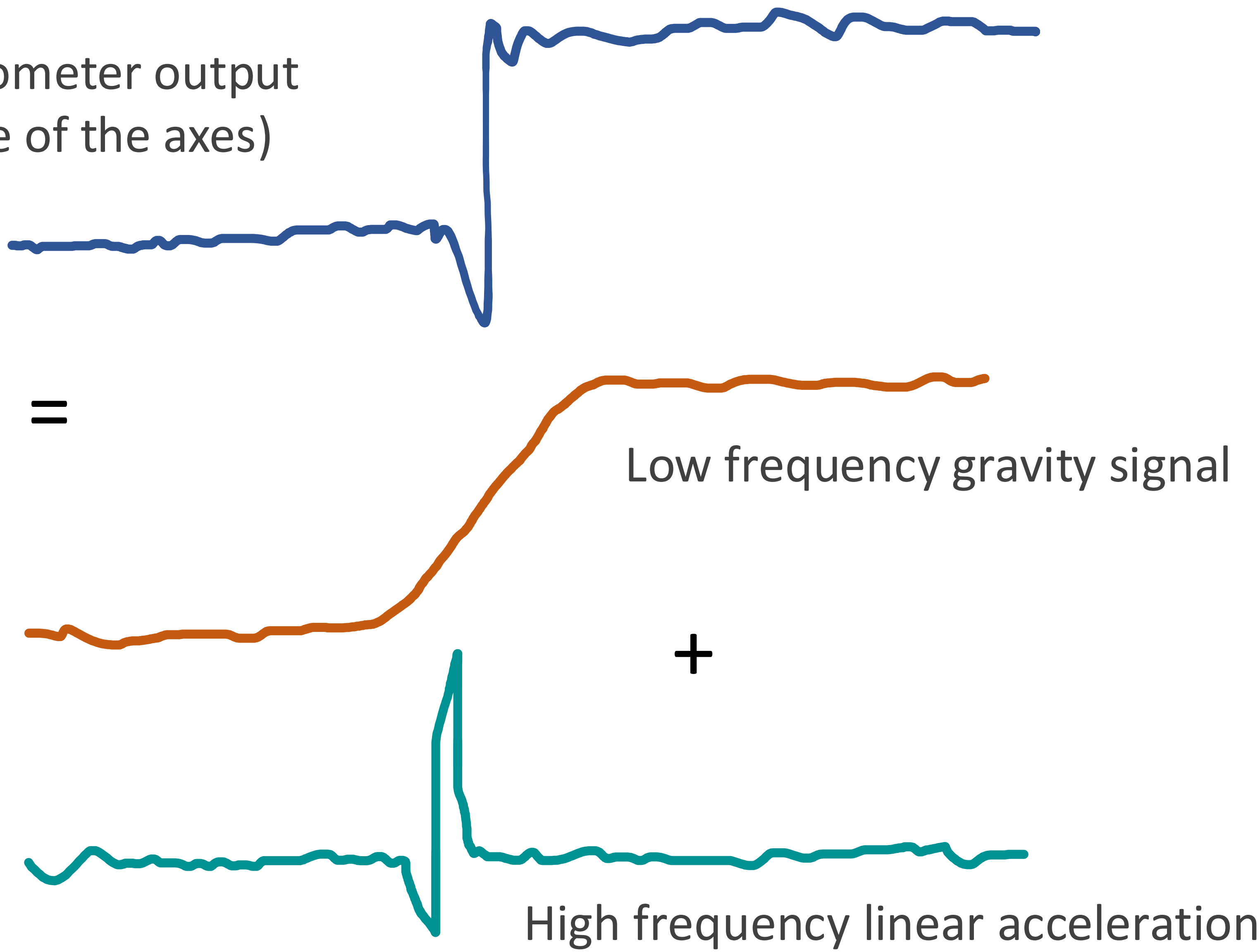
Accelerometer

Gravity filtering



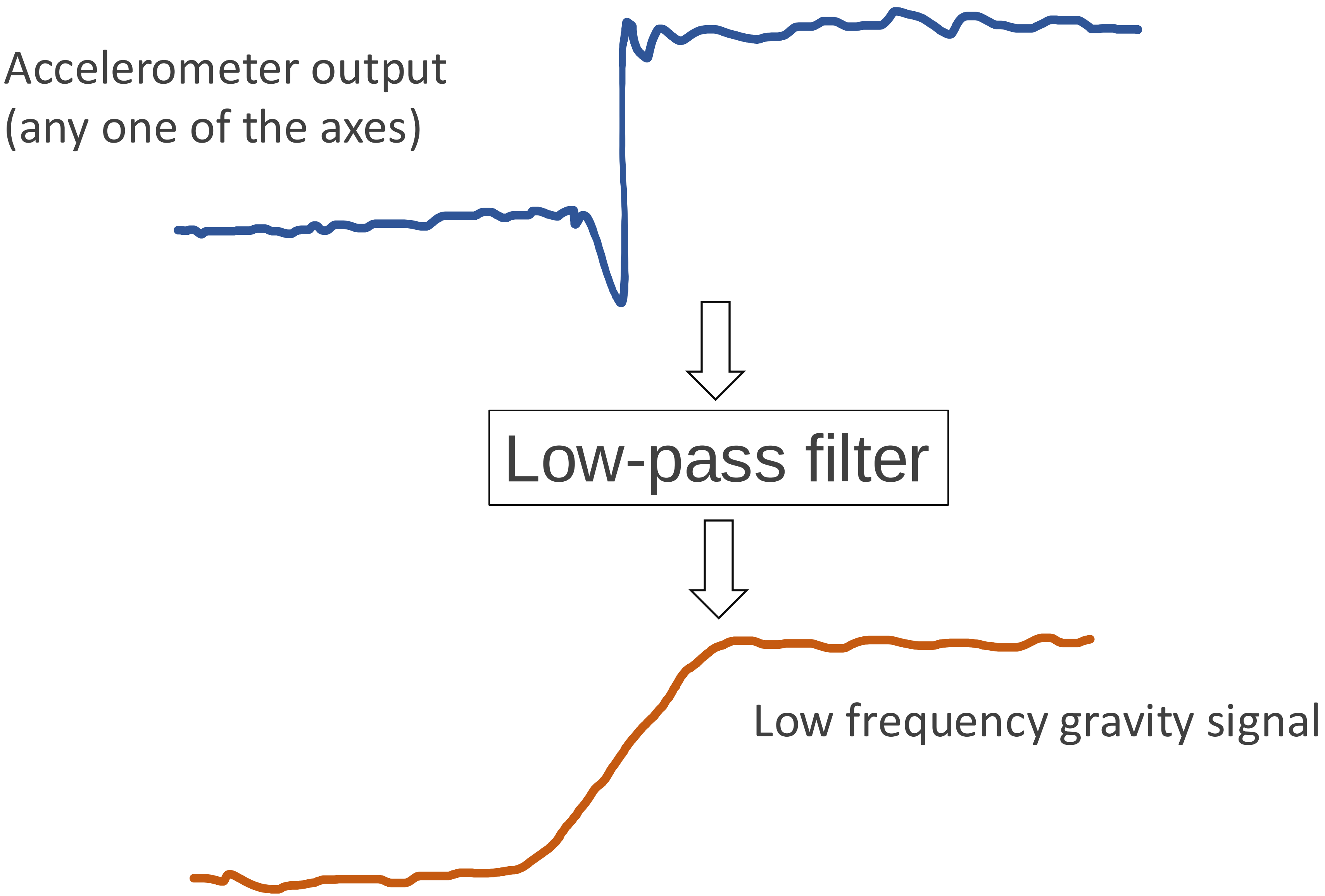
Output = linear acceleration + gravity

Accelerometer output
(any one of the axes)



Accelerometer

Gravity filtering

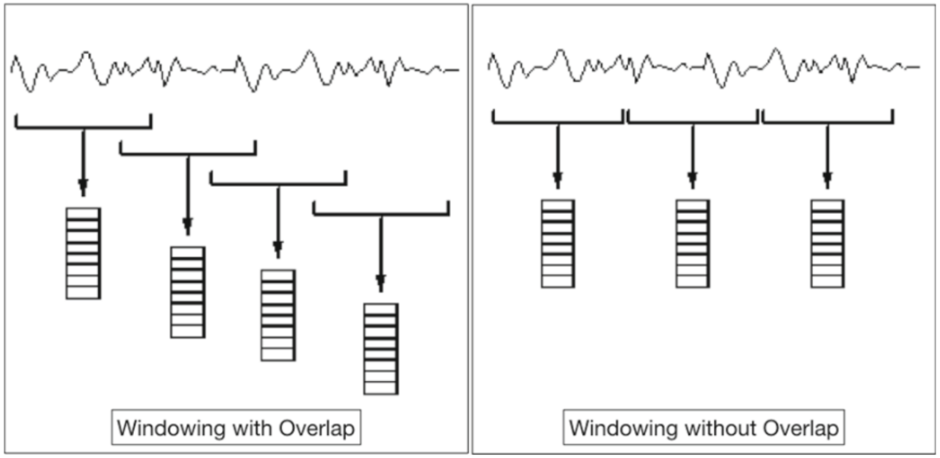


HAR – Data segmentation

Segmentation

Data **segmentation** partitions signals into smaller data segments, also called windows:

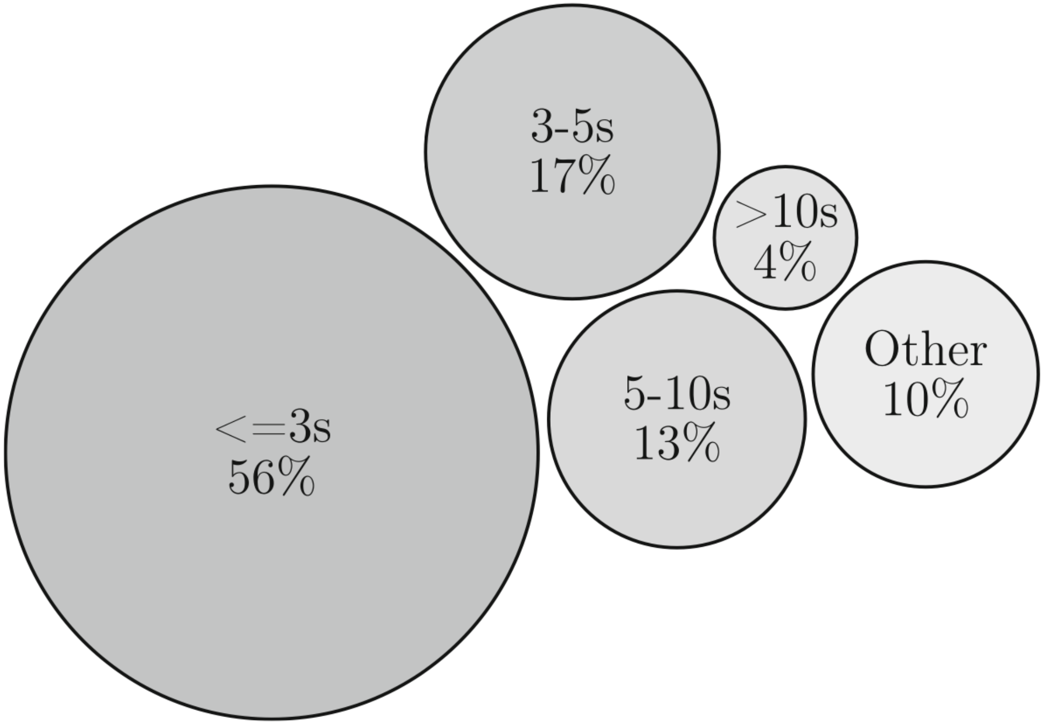
- In **activity-defined** windowing, the initial and end points of each window are selected by detecting patterns of activity changes.
- In **event-defined** windowing, the window is created around a detected event. In some studies, it is also mentioned as windows around the peak.
- In **sliding windowing**, data are split into windows of fixed size, without the gap between two consecutive windows, and, in some cases, overlapping.



HAR – Data segmentation

Segmentation

State-of-the-art **Sliding Window's Size**

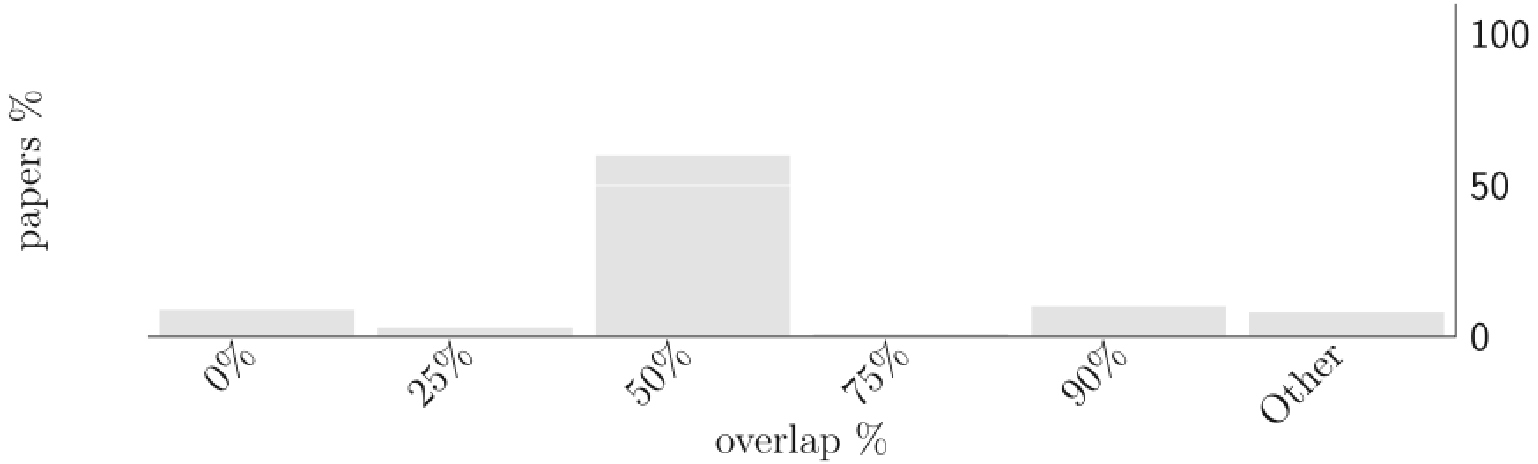


HAR – Data segmentation

Segmentation

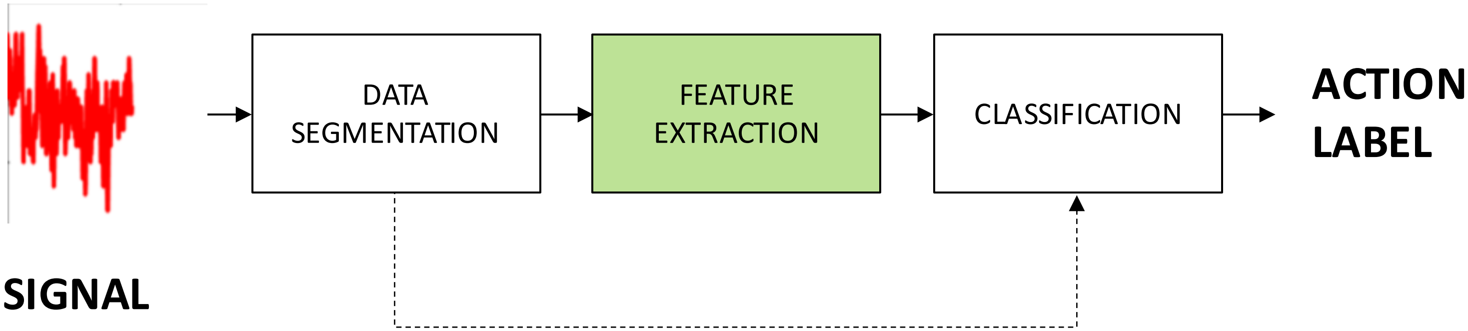
Sliding windows are often **overlapped**, which means that a percentage of a window is repeated in the subsequent window:

1. it avoids noise due to the truncation of data during the windowing process,
2. increases the performance by increasing the data points number.

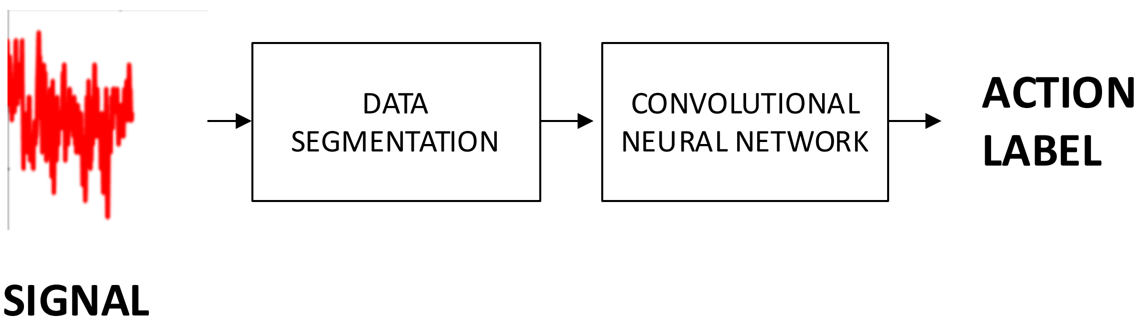


HAR – Pipeline

Methods for recognition



Classifiers: k Nearest Neighbour (k-NN) and Support Vector Machines (SVM) classifiers have been used.



Convolutional Neural Networks: AlexNet, Inception, GoogleNet, ResNet.

* See **additional materials** on Ferrari, A., Micucci, D., Mobilio, M., & Napoletano, P. (2019, June). Hand-crafted features vs residual networks for human activities recognition using accelerometer. In *2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT)* (pp. 153-156). IEEE..

HAR – Feature Extraction

Hand-crafted

Features extraction reduces the data dimensionality while extracting the most important peculiarity of the signal by abstracting each data segment into a high-level representation of the same segment.

Frequency domain features

Feature name	Formula	Description
Entropy	$H(x) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$	Normalized information entropy of the discrete FFT components
Sum of the spectral power components	$ID = x_{0.75} - x_{0.25}$	Difference between third and first quartile of a given segment in each dimension
Mean of the spectral components	$\mu_f = \frac{1}{n} \sum_{j=1}^{n} FFT_j$	Mean of FFT distribution
Median of the spectral components	$Me_f = FFT_{0.5} : F(FFT_{0.5}) = 0.5$	Median of FFT distribution
First cepstral coefficient	$c(1) = \mathcal{F}^{-1}\{\log \lvert FFT(f) \rvert\}$	First coefficient of the cepstrum transformation

HAR – Feature Extraction

Hand-crafted

Time domain features Feature name	Formula	Description
Minimum	$\min_{j=1,\ldots n}(x_j)$	Minimum value of a given segment in each dimension
Maximum	$\max_{j=1,\ldots n}(x_j)$	Maximum value of a given segment in each dimension
Mean	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	Mean value of a given segment in each dimension
Median	$Me = x_{0.5} : F(x_{0.5}) \leq 0.5$	Median value of a given segment in each dimension
Standard Deviation	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$	Standard deviation of a given segment in each dimension
Variance	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	Variance of a given segment in each dimension
Interquartile Difference	$ID = x_{0.75} - x_{0.25}$	Difference between third and first quartile of a given segment in each dimension
Skewness	$skw = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}$	Skewness value of a given segment in each dimension
Kurtosis	$kurt = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4}$	Kurtosis value of a given segment in each dimension
Root mean square	$rms = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$	Root mean square value of a given segment in each dimension
Total Sum	$ts = \sum_{i=1}^{n} x_i$	Total sum value of a given segment in each dimension
Range	$R = max - min$	Range of a given segment in each dimension
Mean of Peak's distance	$m_p = \frac{1}{s^2}\sum_{j=1}^{s}\sum_{i=1}^{s} d(p_i, p_j)$	Mean of distance between peaks of a given segment in each dimension
Fourth central moment	$m_4 = \frac{1}{n}\sum_{j=1}^{n}(x - \bar{x})^4$	Fourth central moment of a given segment in each dimension
Fifth central moment	$m_5 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^5$	Fifth central moment of a given segment in each dimension

HAR – Learned Features

Learned

The goal of feature learning is to automatically discover meaningful representations of raw data to be analyzed

- **Codebooks** considers each sensor data window as a sequence, from which subsequences are extracted and grouped into clusters. Each cluster center is a codeword. Then, each sequence is encoded using a bag-of-words approach using codewords as features.

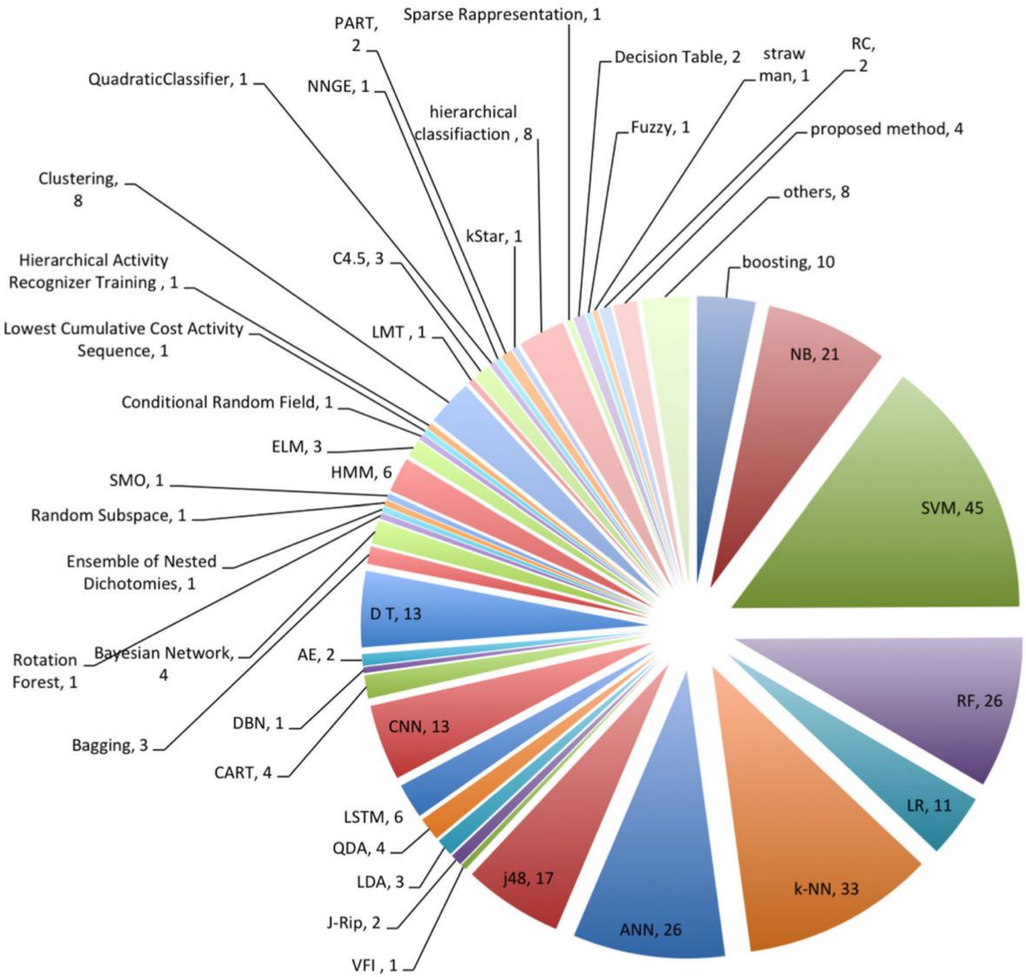
- **Principal Component Analysis** (PCA) is a multi- variate technique, commonly used for dimensionality reduction. The main goal of PCA is the extraction of a set of orthogonal features, called principal component, which are linear combination of the original data and such as the variance extracted from the data is maximal. It is also used for features selection.

- **Deep Learning** uses Neural Networks engines to learn patterns from data

HAR – Classifiers

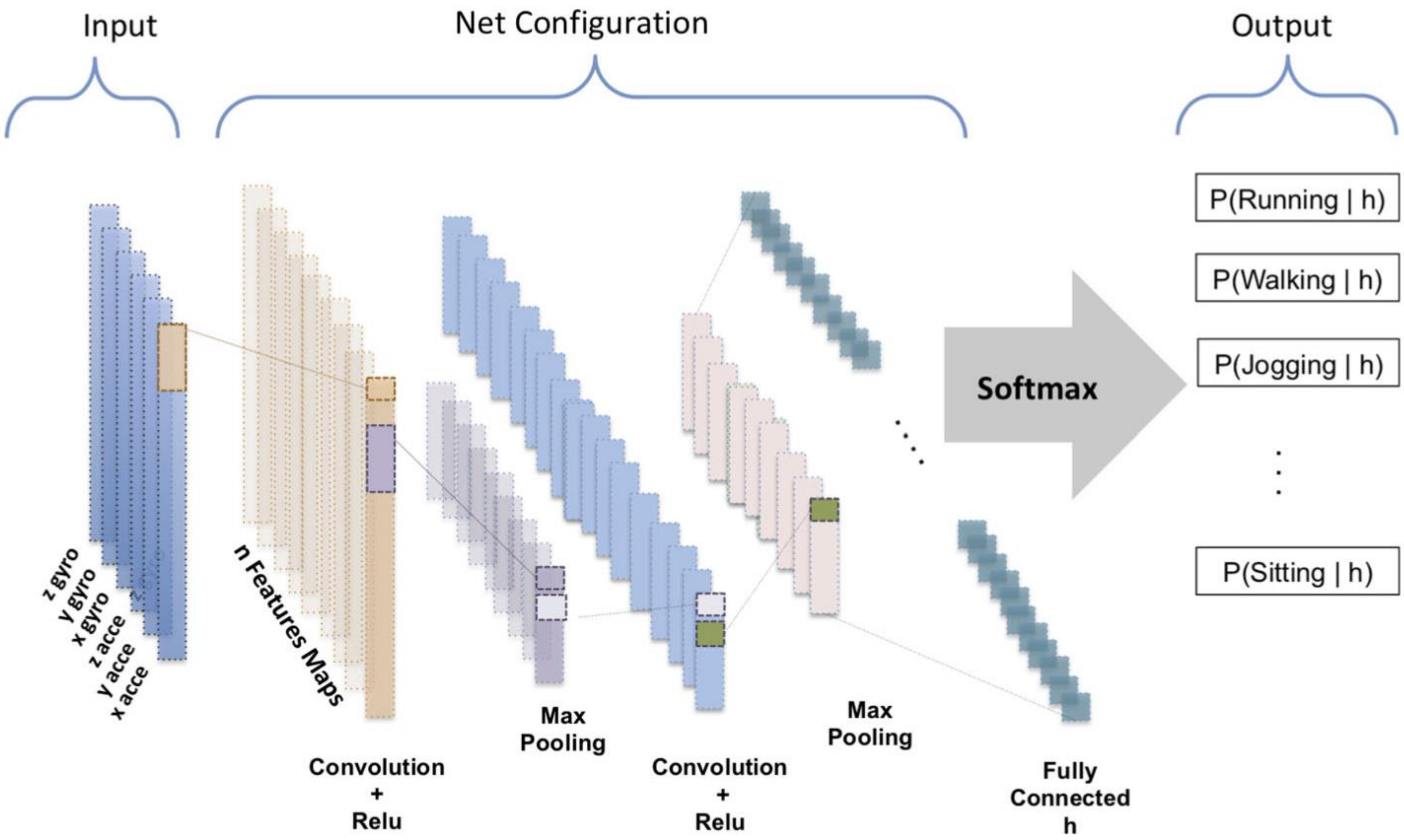
Traditional classifiers

Traditional machine learning and deep learning classifiers distribution



Distance	Formula
Euclidean	$\sqrt{\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{x}_j)^2}$
City Block	$\sum_{i=1}^{n} \lvert\mathbf{x}_i - \mathbf{x}_j\rvert$
Chebychev	$\max_{i=1\ldots n} \lvert\mathbf{x}_i - \mathbf{x}_j\rvert$
Cosine	$1 - \frac{\mathbf{x}_i \mathbf{x}_j^T}{\sqrt{(\mathbf{x}_i \mathbf{x}_i^T)(\mathbf{x}_j \mathbf{x}_j^T)}}$
Correlation	$1 - \frac{(\mathbf{x}_i - \bar{\mathbf{x}}_i)(\mathbf{x}_j - \bar{\mathbf{x}}_j)^T}{\sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_i)(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T}\sqrt{(\mathbf{x}_j - \bar{\mathbf{x}}_j)(\mathbf{x}_j - \bar{\mathbf{x}}_j)^T}}$
Mahalnobis	$\sqrt{(\mathbf{x}_i - \mathbf{x}_j)C^{-1}(\mathbf{x}_i - \mathbf{x}_j)^T}$ where C is the covariance matrix

Traditional machine learning and deep learning classifiers distribution



HAR – Data

Available datasets

Traditional machine learning and deep learning classifiers distribution

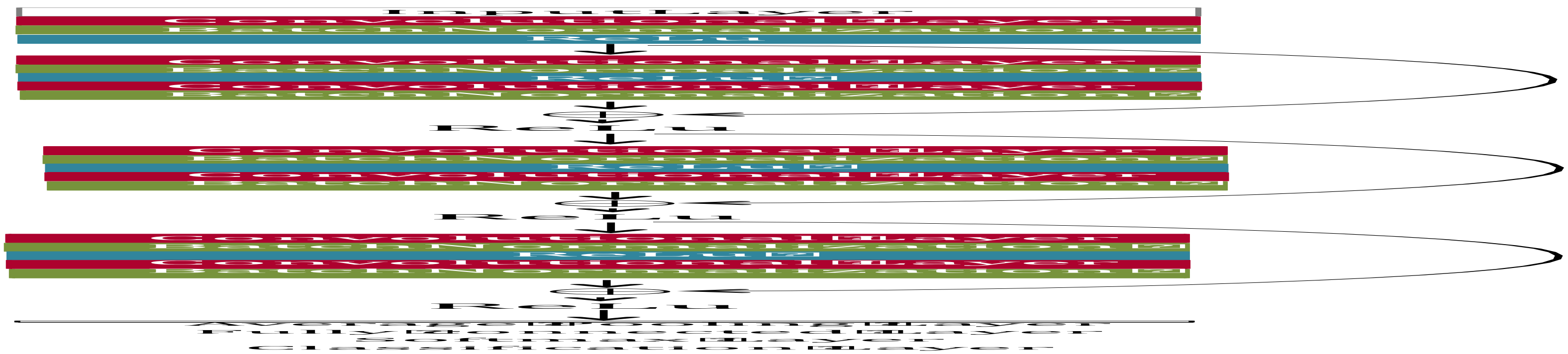
ID	Dataset	# Activity	# Subject	# Devices	Sensors	Sampling rate (Hz)	Metadata
D01	UCI HAR	6 ADL	30	SP(1)	A,G	50	No
D02	Smartphone-based recognition of human Activities and postural transitions data set	6 ADL	30	SP(1)	A,G	50	No
D03	HHAR	6 ADL	9	SP(8),SW(4)	A,G	H	No
D04	Physical activity recognition dataset using Smartphone sensors	6 ADL	4	SP(4)	A,G,M	50	No
D05	Sensors activity dataset	7 ADL	10	SP(5)	AG,M,LA	50	No
D06	Complex human activities dataset	13 ADL	10	SP(2)	A,G,LA	50	No
D07	Motions sense	6 ADL	24	SP(1)	A,G,AT	50	Gender, Age Height,Weight
D08	MobiAct	11 ADL, 4 F	67	SP(1)	A,G,OR	87	Gender, Age Height,Weight
D09	UniMiB-SHAR	9 ADL, 8 F	30	SP(1)	A	50	Gender, Age Height,Weight
D10	UMAFall	12 ADL, 3 F	19	SP(1),IMUs(4)	A,G,M	200,20	Gender, Age Height,Weight
D11	Real world	8 ADL	15	SP(6),SW(1)	A,G,GPS,L,M,S	50	Gender, Age Height,Weight
D12	WISDM	6 ADL	29	SP (1)	A	20	No
D013	Smartphone dataset for HAR in Ambient assisted living (AAL) data Set	6 ADL	30	SP(1)	A,G	50	No
D14	Daily activity dataset	5 ADL	8	SP (1)	A	40	No
D15	HASC2010	6 ADL	96	SP(1)	A	[10-100]	Gender,Height Weight,Shoes Floor,Place
D016	Extrasensory dataset	7 ADL + 109 Specific activities	60	SP(1), SW	A,G,M,CO,LO,S,SM,ST	40,25	No

HAR – Traditional Classifiers vs DNNs

Comparison

In the recent literature, deep learning methods are predominant

Proposed ResNet: The network architecture is made of an initial convolutional block, 3 residual stages, each containing a variable number ***n*** of residual blockoling layer, fully connected layer, and softmax layer.



Layer name	shape
conv1	$\{1{\times}3\} \times n$
conv2_n	$\{1{\times}3 \times f_{maps}\} \times n$
conv3_n	$\{1{\times}3 \times 2f_{maps}\} \times n$
conv4_n	$\{1{\times}3 \times 4f_{maps}\} \times n$
avg_pool_x	$1{\times}32$
fully conn.	$(1 \times 4f_{maps}) \times 15$
softmax	$1{\times}15$

For each dataset, the best values for n and f_{maps} have been found by following a grid search approach: n ranged between 3 and 21, while fmaps ranged between 10 and 200.

* See **additional materials** on Hand-crafted Features vs Residual Networks for Human Activities Recognition using Accelerometer (Anna Ferrari, Daniela Micucci, Marco Mobilio, Paolo Napoletano) In 2019 IEEE 23RD International Symposium on Consumer Technologies
Human Activities Recognition using Accelerometer and Gyroscope (Anna Ferrari, Daniela Micucci, Marco Mobilio, Paolo Napoletano) In 2019 European Conference on Ambient Intelligence

HAR – Traditional Classifiers vs DNNs

Comparison

- **Raw data** (denoted as *raw*): x,y, and z accelerometer segments (without any kind of processing) are concatenated and used as feature vectors;
- **Magnitude** of the segments (denoted as *magn*);
- **21 features** extracted from the magnitude of the segments (denoted as *hc magn*).
- **21 features** extracted from each of the three segments along the three axes x, y, and z (denoted as *hc raw*). The total number of features is 63.

Minimum	$min = \min_{j=1,\dots,n}(x_j)$
Maximum	$max = \max_{j=1,\dots,n}(x_j)$
Mean	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Median	$Me = x_{0.5} : F(x_{0.5}) = 0.5$
Standard Deviation (SD)	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$
Variance	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$
Fourth central moment	$m_4 = \frac{1}{n}\sum_{j=1}^{n}(x-\bar{x})^4$
Fifth central moment	$m_5 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^5$
Skewness	$S = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{s^3}$
Kurtosis	$K = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{s^4}$
Root Mean Square (RMS)	$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$
Interquartile Difference	$ID = x_{0.75} - x_{0.25}$
Total Sum	$TS = \sum_{i=1}^{n} x_i$
Range	$R = max - min$
Entropy	$H(x) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$
SD of the intervals between two successive peaks	$SDNN = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(PP_j-\overline{PP})^2}$
RMS of the differences between two successive peaks	$RMSSD = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n-1}(PP_{j+1}-PP_j)^2}$
Number of pairs of successive peaks intervals that differ by more than 50 ms	$pNN50 = p(\lvert PP_{j+1}-PP_j\rvert > 50)$
Sum of the spectral power components	$SP = \frac{1}{n}\sum_{j=1}^{f}\lvert FFT_j\rvert^2$
Mean of the spectral components	$\mu_f = \frac{1}{n}\sum_{j=1}^{n} FFT_j$
Median of the spectral components	$Me_f = FFT_{0.5} : F(f_{0.5}) = 0.5$

* See **additional materials** on Hand-crafted Features vs Residual Networks for Human Activities Recognition using Accelerometer (Anna Ferrari, Daniela Micucci, Marco Mobilio, Paolo Napoletano) In 2019 IEEE 23RD International Symposium on Consumer Technologies
Human Activities Recognition using Accelerometer and Gyroscope (Anna Ferrari, Daniela Micucci, Marco Mobilio, Paolo Napoletano) In 2019 European Conference on Ambient Intelligence

HAR – Traditional Classifiers vs DNNs

Comparison

TABLE IV
EXPERIMENTAL RESULTS - MEAN CLASS ACCURACY(STANDARD DEVIATION CLASS ACCURACY): SVM VS RESNET

	SVM				ResNet
Dataset	raw	magn	hc raw	hc magn	
UCI-HAR	79.51 (± 17.40)	53.10 (± 25.48)	79.47 (± 20.59)	48.45 (± 22.12)	**90.73 (± 10.92)**
MobiAct	77.93 (± 22.71)	63.63 (± 24.13)	76.73 (± 26.11)	59.95 (± 23.94)	**92.98 (± 8.65)**
MotionSense	90.04 (± 14.36)	78.22 (± 29.59)	96.39 (± 3.79)	83.45 (± 21.13)	**99.47 (± 0.87)**
UniMiB-SHAR	58.26 (± 16.85)	52.27 (± 18.10)	58.08 (± 16.70)	50.81 (± 15.49)	**88.59 (± 8.52)**

TABLE V
EXPERIMENTAL RESULTS - MEAN CLASS ACCURACY(STANDARD DEVIATION CLASS ACCURACY): k-NN VS RESNET

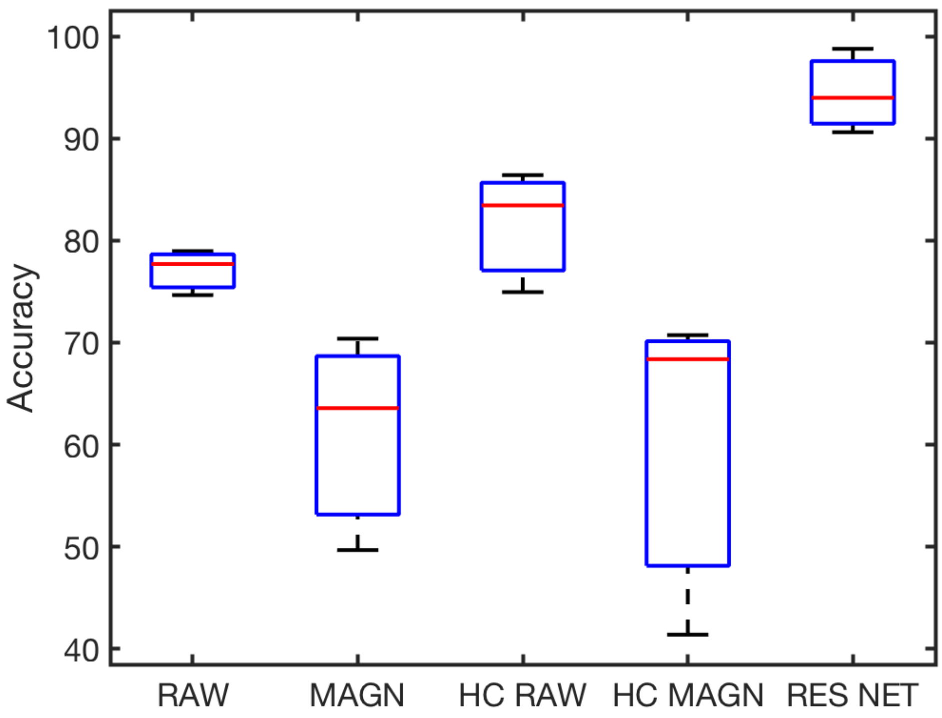
	k-NN				ResNet
Dataset	raw	magn	hc raw	hc magn	
UCI-HAR	73.71 (± 26.78)	46.92 (± 29.89)	69.35 (± 17.04)	37.75 (± 13.39)	**90.73 (± 10.92)**
MobiAct	87.69 (± 9.07)	77.81 (± 13.60)	91.86 (± 6.72)	80.50 (± 10.74)	**92.98 (± 8.65)**
MotionSense	79.19 (± 31.83)	73.51 (± 25.16)	95.82 (± 5.61)	81.34 (± 20.30)	**99.47 (± 0.87)**
UniMiB-SHAR	61.97 (± 11.83)	55.13 (± 14.29)	65.74 (± 12.99)	52.22 (± 11.70)	**88.59 (± 8.52)**

The average gap between hand-crafted features combined with traditional classifiers and deep learning is about **15%**

* See **additional materials** on Hand-crafted Features vs Residual Networks for Human Activities Recognition using Accelerometer (Anna Ferrari, Daniela Micucci, Marco Mobilio, Paolo Napoletano) In 2019 IEEE 23RD International Symposium on Consumer Technologies
Human Activities Recognition using Accelerometer and Gyroscope (Anna Ferrari, Daniela Micucci, Marco Mobilio, Paolo Napoletano) In 2019 European Conference on Ambient Intelligence

HAR – Traditional Classifiers vs DNNs

Comparison



- ResNet is the best performing with an average accuracy across datasets of about **93**%
- The second best across classifiers and datasets are the *hc raw* features with an average accuracy of about 80 %.
- The third best are the *raw* features with an average accuracy of about 76%.
- The worst are the *magnitude* and *magnitude raw* features with an average accuracy of about 62%.

* See **additional materials** on Hand-crafted Features vs Residual Networks for Human Activities Recognition using Accelerometer (Anna Ferrari, Daniela Micucci, Marco Mobilio, Paolo Napoletano) In 2019 IEEE 23RD International Symposium on Consumer Technologies
Human Activities Recognition using Accelerometer and Gyroscope (Anna Ferrari, Daniela Micucci, Marco Mobilio, Paolo Napoletano) In 2019 European Conference on Ambient Intelligence

HAR – Traditional Classifiers vs DNNs

Comparison

According to recent studies, the combination of accelerometer and gyroscope signals, also called multimodal recognition, increases the accuracy in HAR with respect to the use of each signal alone.

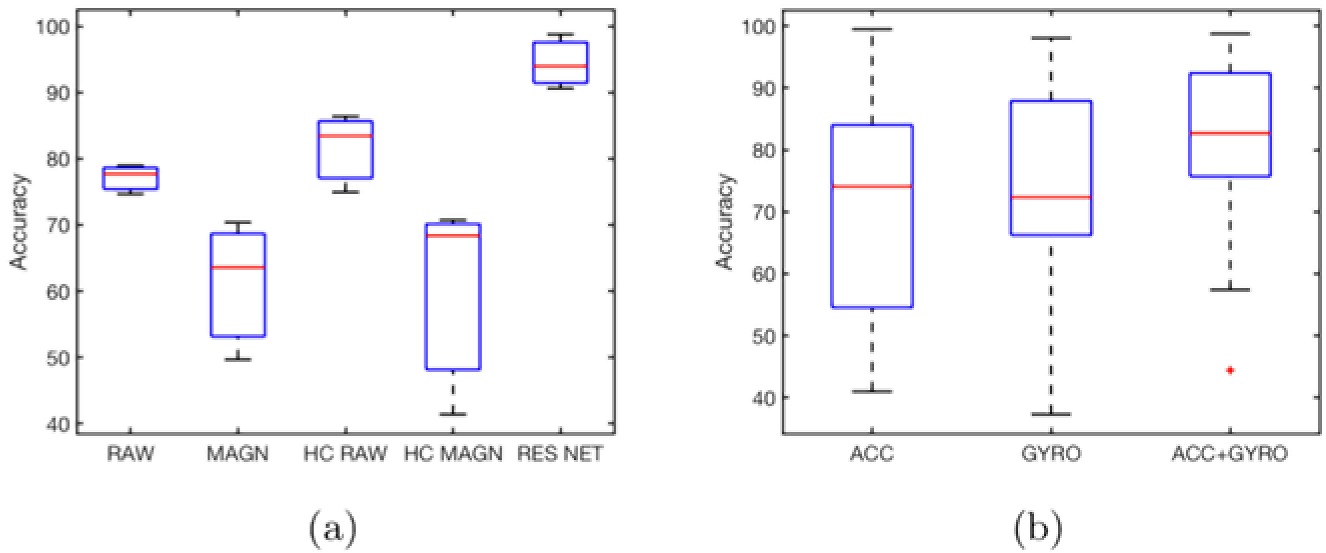


Fig. 1. Experiments. (a) comparison across datasets between hand-crafted and ResNet. (b) comparison across datasets and methods between multimodality and single modality.

* See **additional materials** on Hand-crafted Features vs Residual Networks for Human Activities Recognition using Accelerometer (Anna Ferrari, Daniela Micucci, Marco Mobilio, Paolo Napoletano) In 2019 IEEE 23RD International Symposium on Consumer Technologies
Human Activities Recognition using Accelerometer and Gyroscope (Anna Ferrari, Daniela Micucci, Marco Mobilio, Paolo Napoletano) In 2019 European Conference on Ambient Intelligence

Arm-gesture recognition

Arm gesture recognition

Inertial sensors can be exploited by machine learning algorithms to perform fine and coarse-grained **human action recognition** with a high level of accuracy. Inertial signals, in contrast with visual signals taken with video cameras, **preserve privacy** of the user thus being a very good candidate for development of applications for the remote control of devices.

Target

- A **dictionary of arm gestures** using inertial sensors embedded in a **smart bracelet** is created.
- A machine learning method for the recognition of arm gestures that is **invariant** with respect to the **time** employed to perform the gesture, to the **subject** who performs the gesture and the position of the arm.

Arm gesture recognition

1. **Gesture Recognition:** The goal of this first task is to recognize the input gesture into one of the N classes of the acquired dataset.

2. **User Identification:** This is the recognition of the identity of the user that performed one of the gestures taken from the vocabulary.

3. **User Authentication:** This is the task of user authentication, meaning the ability to verify the identity of a user attempting to use the wristband.

4. **Gesture Authentication:** This is the task of gesture authentication, meaning the ability to verify that a given type of action has been performed whatever is the user.

* See **additional materials** on Bianco, S., Napoletano, P., Raimondi, A., & Rima, M. (2022). U-wear: User recognition on wearable devices through arm gesture. *IEEE transactions on human-machine systems*, *52*(4), 713-724.

Arm gesture recognition

User Authentication: The third experiment is focused on the task of user authentication, meaning the ability to verify the identity of a user attempting to use the wristband.

- *Gesture independent*, where the aim is to verify the identity of a user whatever is the arm gesture performed. For example, being able to verify the identity of user X performing the gesture A as well as the gesture B, C, etc.

- *Gesture dependent*, where the aim is to verify the identity of a user performing a given arm gesture. For example, being able to verify the identity of user X performing the gesture A and not B, C, etc.

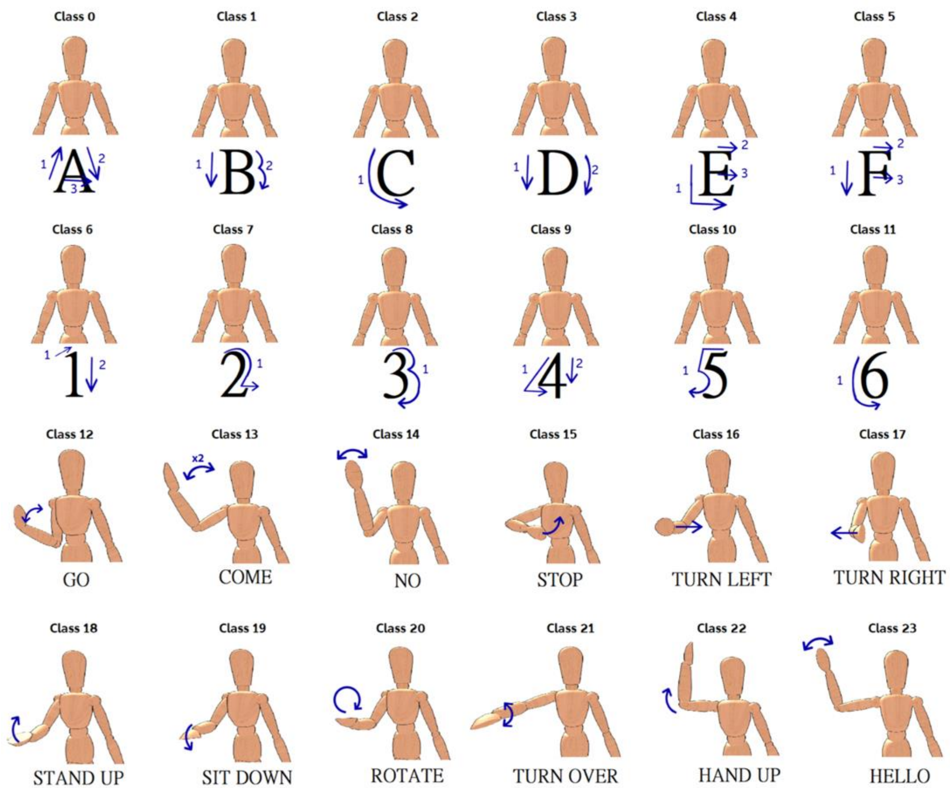
* See **additional materials** on Bianco, S., Napoletano, P., Raimondi, A., & Rima, M. (2022). U-wear: User recognition on wearable devices through arm gesture. *IEEE transactions on human-machine systems*, *52*(4), 713-724.

Arm gesture dictionary

Taking inspiration from [1] and [2] , a combination of gestures and symbols is proposed. To these "no action" class is added which is performed by randomly moving the arm or by keeping the arm steady.

A total of 25 classes (12 symbols + 12 gestures + 1 no-actions)
Taken from more than 30 people which randomly performed each gesture 5 times

http://www.ivl.disco.unimib.it/activities/u-wear/

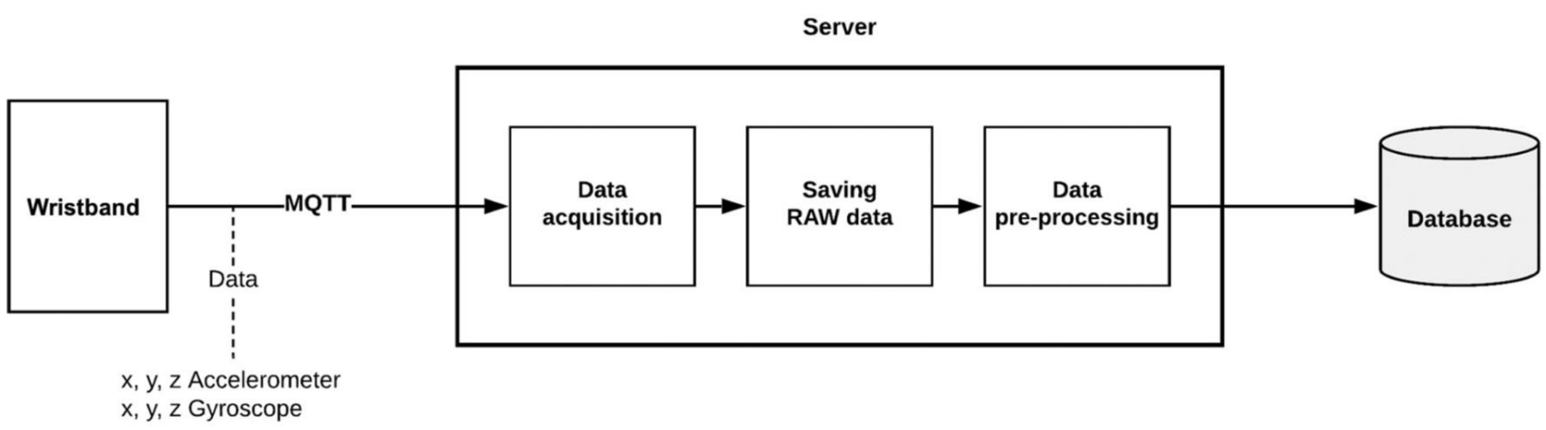


[1] Li, C., Xie, C., Zhang, B., Chen, C., & Han, J. (2018). Deep Fisher discriminant learning for mobile hand gesture recognition. *Pattern Recognition*, *77*, 276-288
[2] Wu, Y., Wu, Z., & Fu, C. (2018). Continuous Arm Gesture Recognition Based on Natural Features and Logistic Regression. IEEE Sensors Journal, 18(19), 8143-8153.

Arm gesture dictionary

Data collection

A **wristband** equipped with a **IMU** is used for data collecting. Data are sent through the **message queue telemetry transport (MQTT) protocol**
from the wristband to a server. On server side, a data buffer is used until the gesture performed by the user is finished. Then, raw data are saved and after a preprocessing step data are also saved into a database.

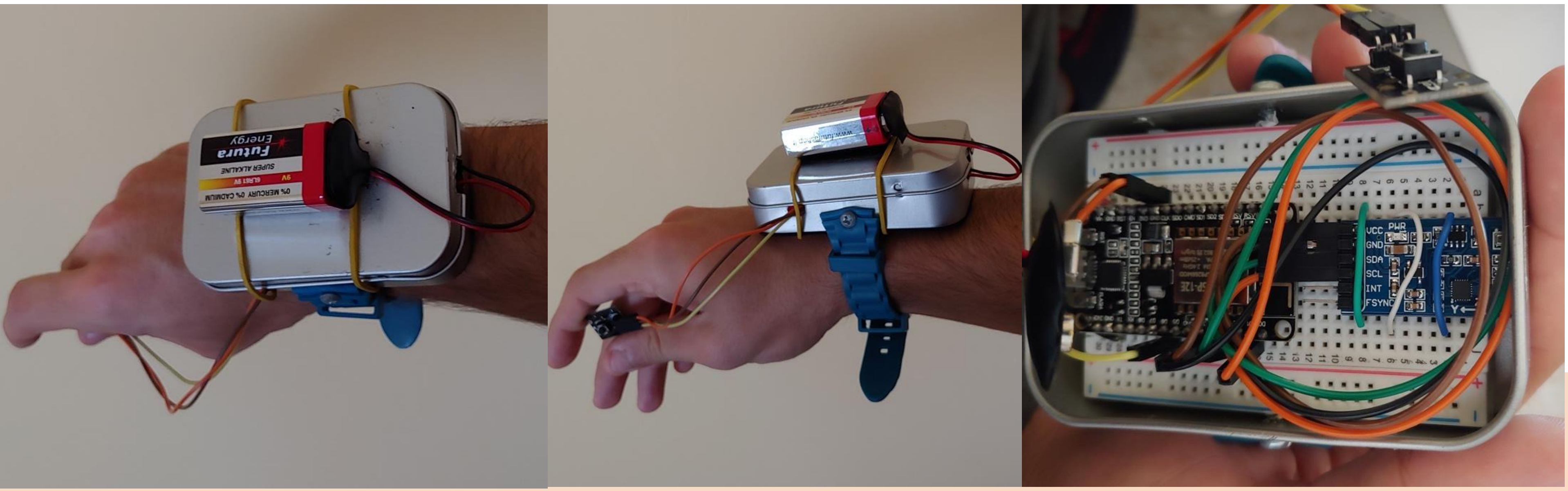


* See **additional materials** on Bianco, S., Napoletano, P., Raimondi, A., & Rima, M. (2022). U-wear: User recognition on wearable devices through arm gesture. *IEEE transactions on human-machine systems*, *52*(4), 713-724.

Edge

A smart bracelet….
Now it is more a handcrafted bracelet than a smart one ☺

It records 3 axial **accelerometer** and **gyroscope**



The bracelet is based on a commercial Micro (wi-fi-enabled) coupled with an IMU sensor. It is subscribed to a **MQTT** topic. It sends data at **200Hz** to a cloud server which performs the classification pipeline

* See additional materials on Bianco, S., Napoletano, P., Raimondi, A., & Rima, M. (2022). U-wear: User recognition on wearable devices through arm gesture. *IEEE transactions on human-machine systems*, *52*(4), 713-724.

Edge

The **wristband** consists of a computing unit, a wireless communication unit, and an inertial sensor.

1. Inertial measurement unit (**IMU**) sensor, including a 3-axis accelerometer and 3-axis gyroscope.
2. **NodeMCU** microcontroller, equipped with the ESP8266 chip.
3. Activation **button** (KY-004 module).
4. External power source.

The magnetometer is not considered to have an initial position not limited by height, position in space or by the inclination of the arm.

The server is a **workstation** equipped with Ubuntu 22.04

https://www.waveshare.com/product/modules/10-dof-imu-sensor-c.htm
https://www.nodemcu.com/index_en.html

10 DOF IMU Sensor (C), Inertial Measurement Unit,

- Power: 3.3V~5.5V (internal voltage regulation with low dropout)
- Accelerometer
  - Resolution: 16 bit
  - Measurement range (configurable): ±2, ±4, ±8, ±16g
  - Operating current: 450uA
- Gyroscope
  - Resolution: 16 bit
  - Measurement range (configurable): ±250, ±500, ±1000, ±2000°/sec
  - Operating current: 3.2mA
- Compass/Magnetometer
  - Resolution: 14bit
  - Measurement range: ±4800μT
  - Operating current: 280uA
- Barometric pressure sensor
  - Barometric resolution: 0.0016hPa
  - Temperature resolution: 0.01°C
  - Measurement range: 300~1100hPa (altitude: +9000m ~ -500m)
  - Barometric relative accuracy (700hPa~900hPa, 25°C~40°C): ±0.12hPa (±1m)
  - Operating current (1Hz update rate, ultra-low power mode): 2.8uA

NodeMCU microcontroller, equipped with the ESP8266 chip

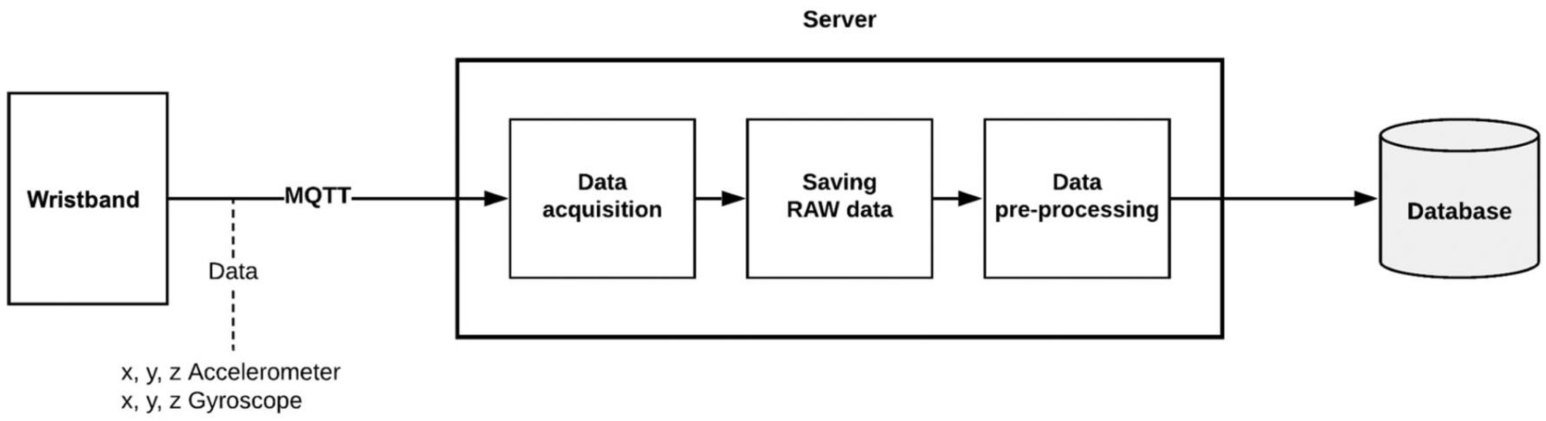
NodeMCU Technical Specifications		
	Official NodeMCU Amica	LoLin NodeMCU
Microcontroller	ESP-8266 32-bit	ESP-8266 32-bit
NodeMCU Model	Amica	Clone LoLin
NodeMCU Size	49mm x 26mm	58mm x 32mm
Pin Spacing	**0.9" (22.86mm)**	1.1" (27.94mm)
Clock Speed	80 MHz	80 MHz
USB to Serial	CP2102	CH340G
USB Connector	Micro USB	Micro USB
Operating Voltage	3.3V	3.3V
Input Voltage	4.5V-10V	4.5V-10V
Flash Memory/SRAM	4 MB / 64 KB	4 MB / 64 KB
Digital I/O Pins	11	11
Analog In Pins	1	1
ADC Range	0-3.3V	0-3.3V
UART/SPI/I2C	1 / 1 / 1	1 / 1 / 1
WiFi Built-In	802.11 b/g/n	802.11 b/g/n
Temperature Range	-40C - 125C	-40C - 125C

Server-side

A **Python-based** application is running on the server. This application is listening on the **MQTT topic** used by the wristband to transmit the IMU sensor data to the server.

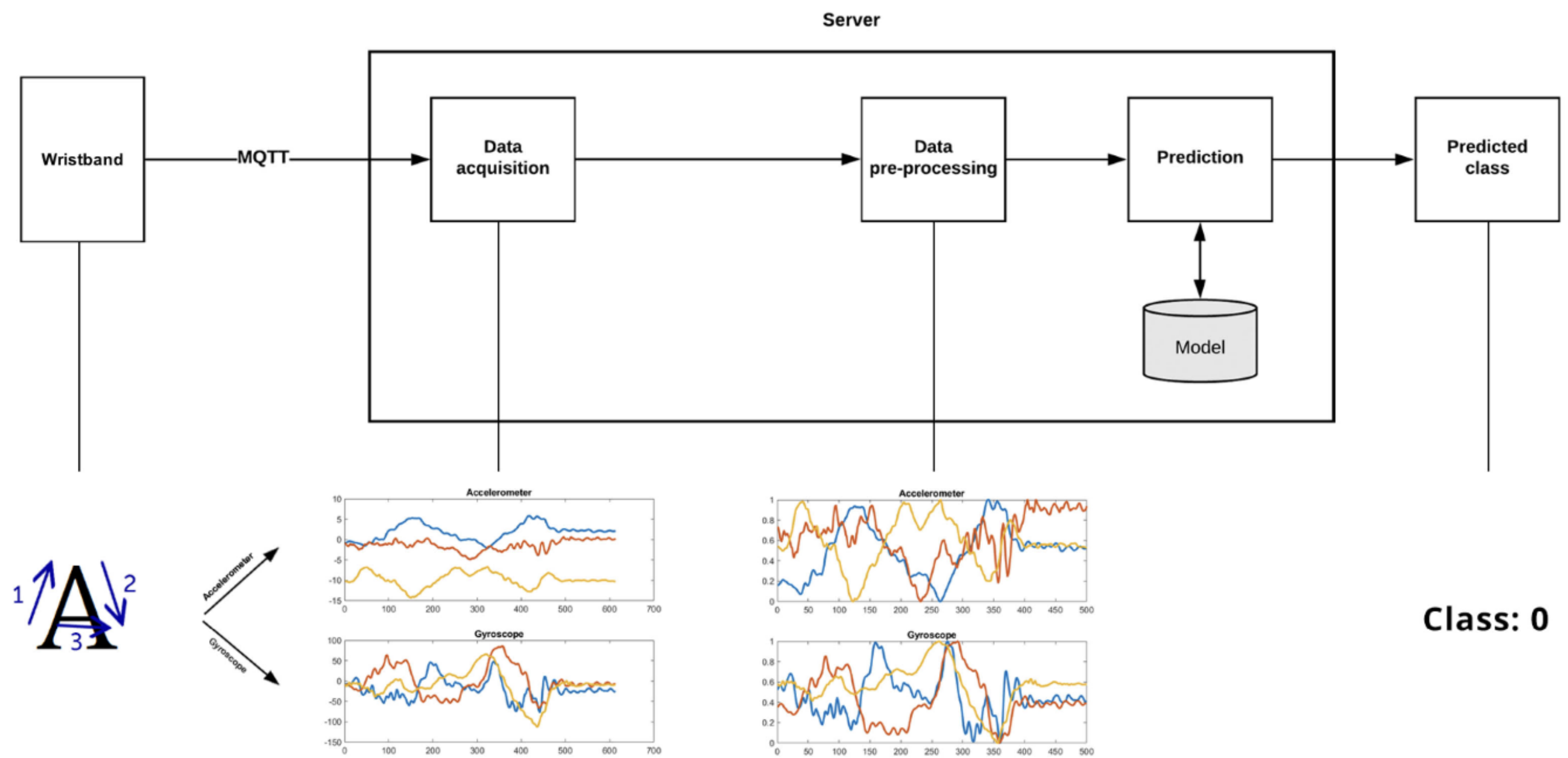
Once the data from the wristband is received, then data are kept in a buffer until the user has his **hand steady** for a given time.

The data are saved in **raw format** and after a **preprocessing** step it is saved to the final **database** or processed for **recognition**.

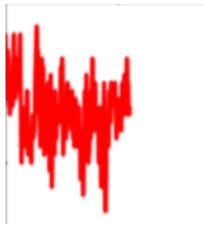


Arm gesture recognition pipeline

Recognition pipeline

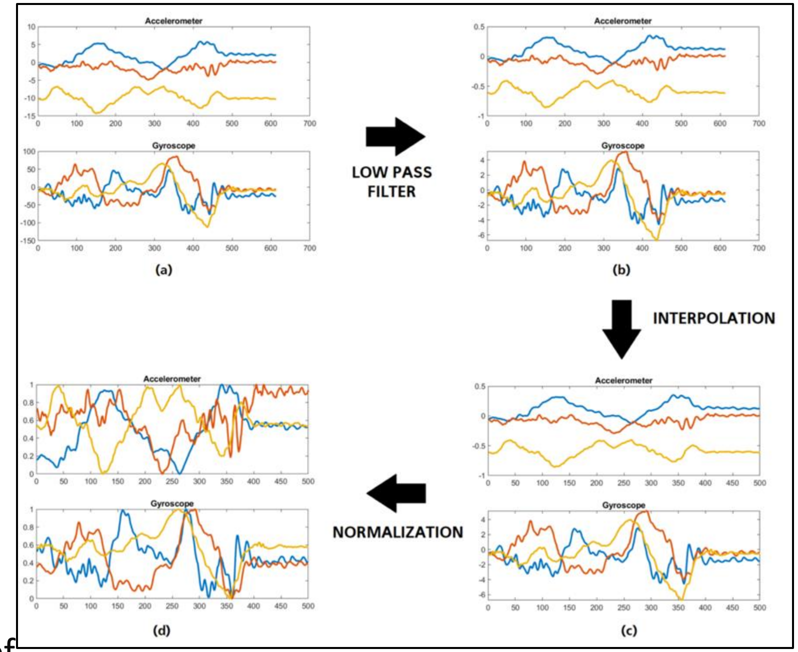


Arm gesture recognition pipeline



SIGNAL

DATA PREPROCESSING



Noise reduction

Each incoming signal is processed with a low pass filter in order to reduce noise components at a higher frequency

Time-length invariance

Recorded data are subsampled or upsampled to a standard length of 500 samples, which is equivalent to a recording of length 2.5 s considering 200 samples per second.

Space-length invariance

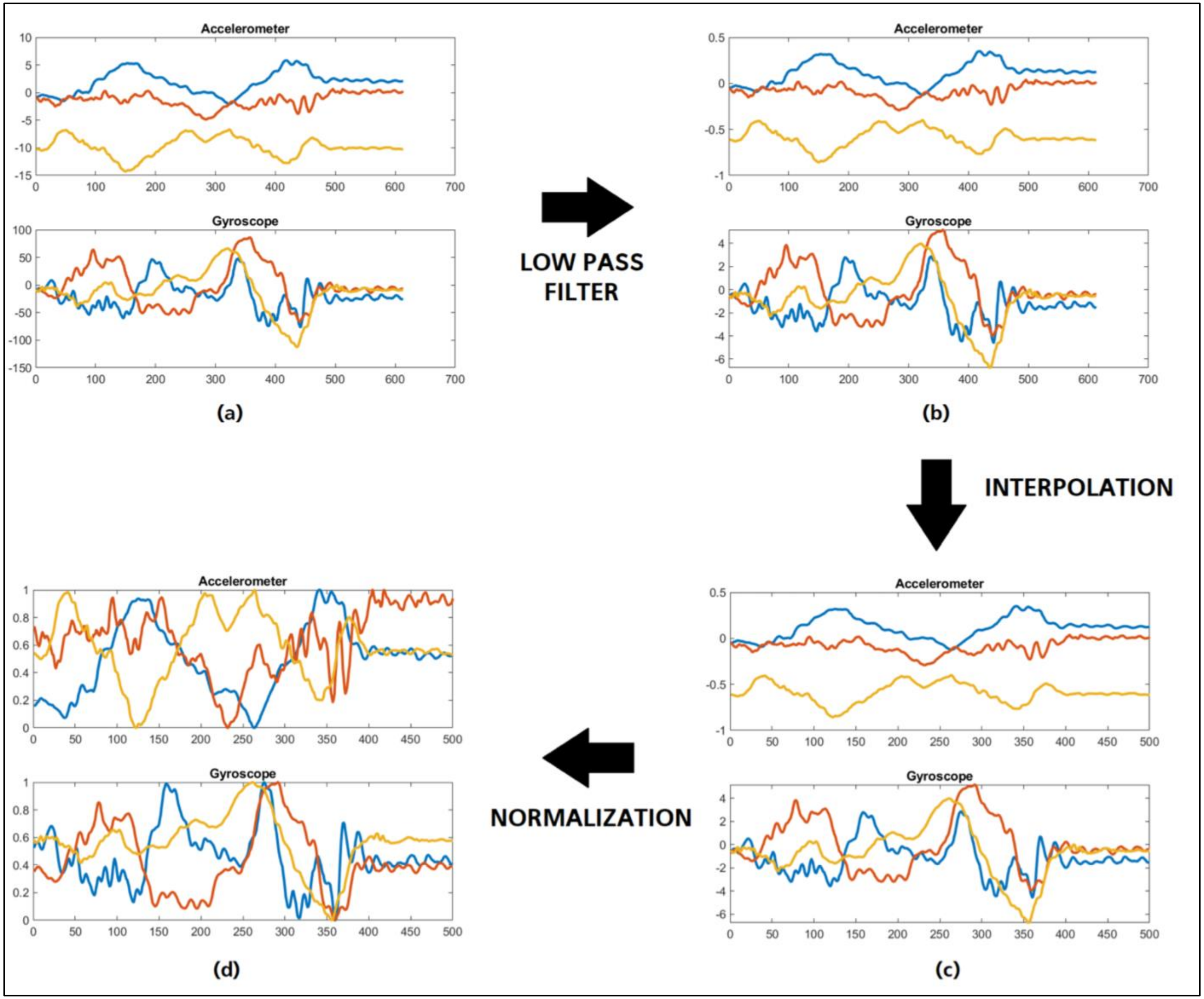
Each gesture can be performed by ranging from a few centimeters to hundreds of centimeters.

Normalization

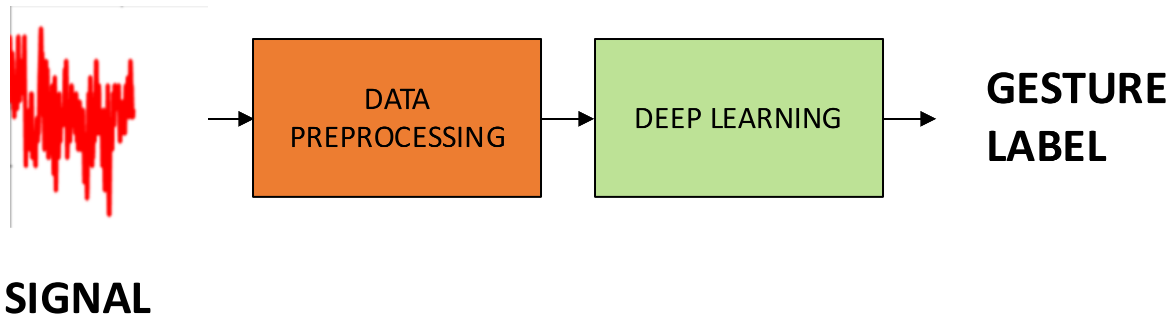
It depends on the recognition task. In the case of gesture recognition, we constrained each recording gesture to a range between 0 and 1. In the case of user identification, we applied the Z-score normalization

Arm gesture recognition pipeline

AI



Arm gesture recognition pipeline



Deep learning

Each sample is of size 500 x 6 (3 accelerometer + 3 gyroscope signals). This is the input of the network (*N*=6 and *T*=500).

$$G = \{G_1, \ldots, G_T\} \qquad G_t = (x_1(t), \ldots, x_N(t))$$

We consider two different networks:

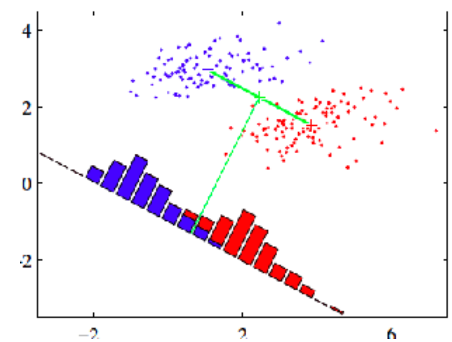
- F-BLSTM (Fisher criterion Bidirectional Long-Short Term Memory)
- F-BGRU (Fisher criterion Bidirectional Gated Recurrent Unit)

Combined Loss

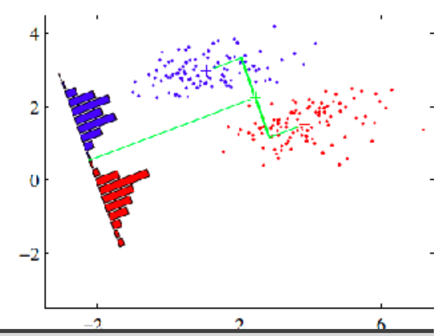
Fisher’s criterion combined with cross entropy:
This criterion maximizes the variance of the class means while minimizes the variance of the individual classes

$$\boxed{\mathcal{L} = \mathcal{L}_s + \theta \mathcal{L}_f}$$

$$\mathcal{L}_s = -\frac{1}{m}\sum_{i=1}^{m} \log \frac{e^{W_{y_i}^T O_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T O_i + b_j}},$$



$$\mathcal{L}_f = \frac{1}{m}\sum_{i=1}^{m} \|O_i - \mu_{y_i}\|_2^2 - \frac{\delta}{n\,(n-1)} \sum_{j=1,k=1}^{n} \|\mu_j - \mu_k\|_2^2$$

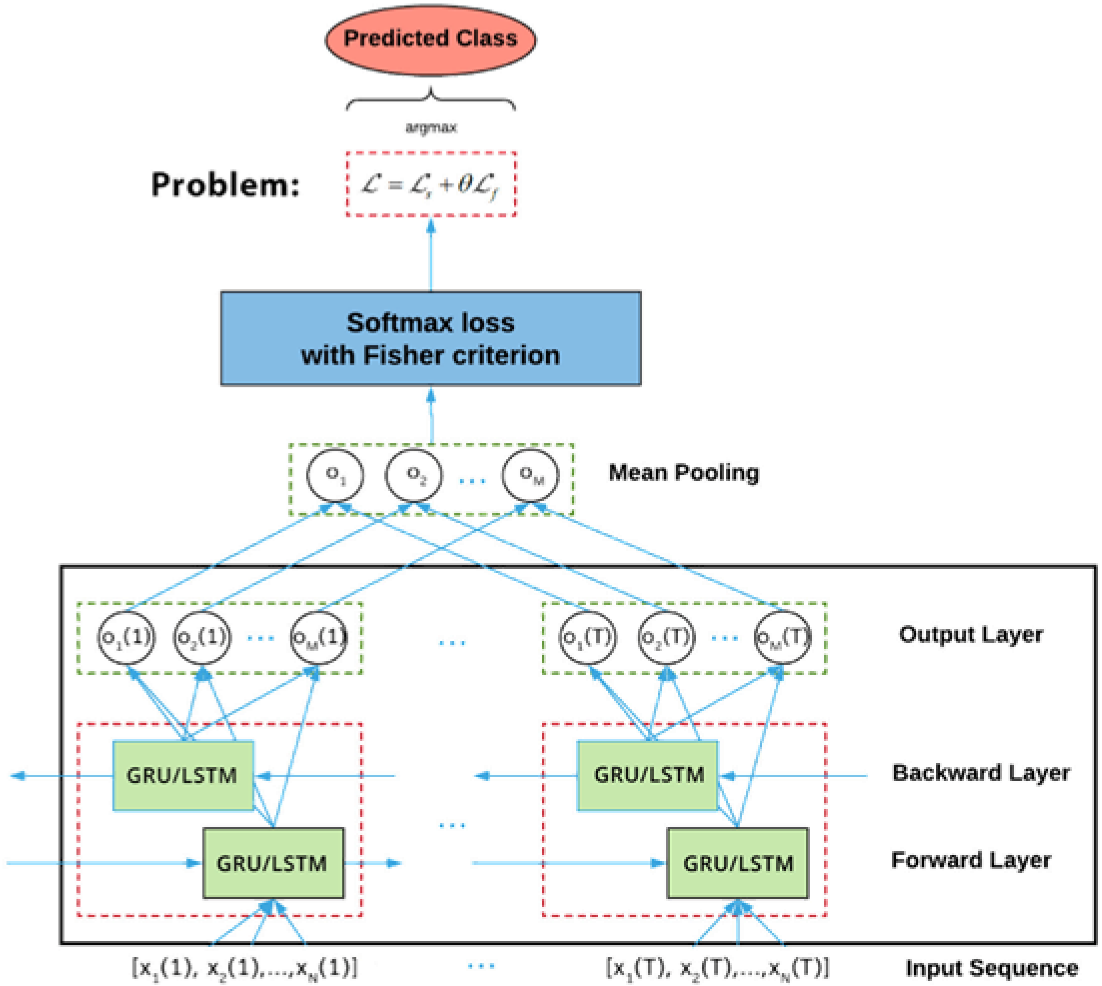


Arm gesture recognition pipeline

Two different networks:

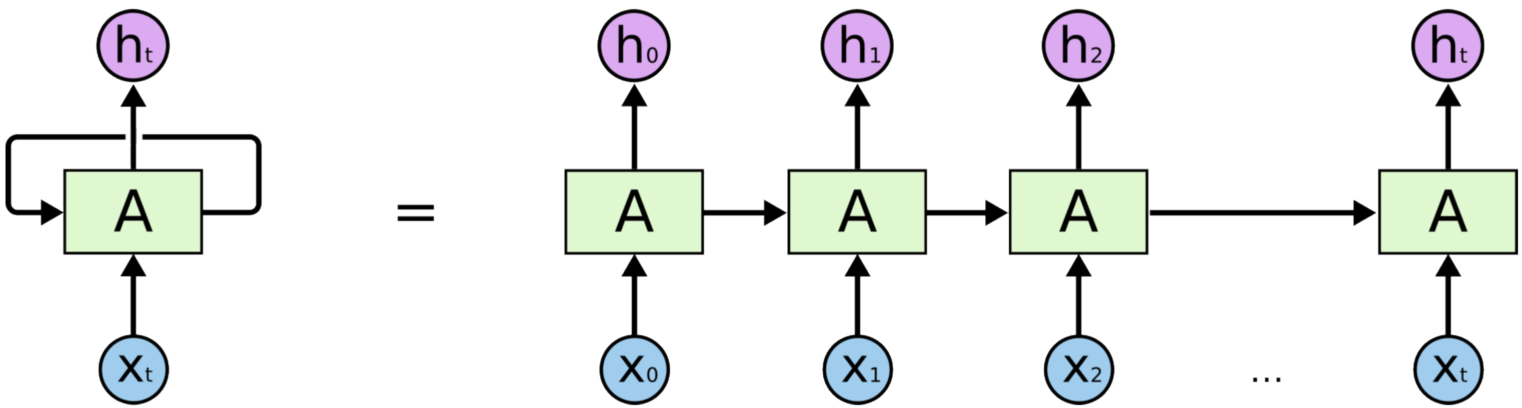
- F-BLSTM (Fisher criterion Bidirectional Long-Short Term Memory)
- F-BGRU (Fisher criterion Bidirectional Gated Recurrent Unit)

- We have M=64 hidden layers that are fully connected to the backwards and forwards LSTM units. N=6 and T=500.



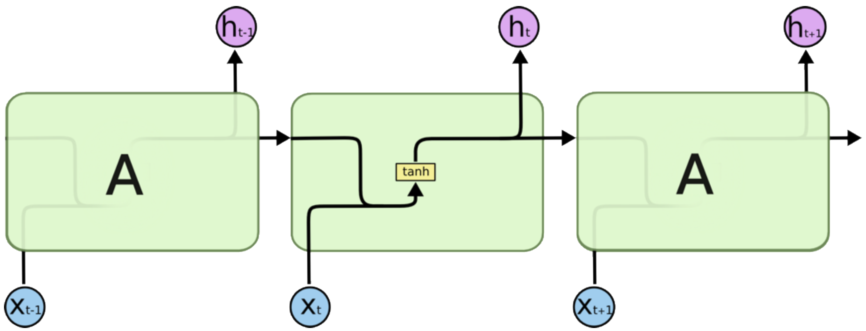
Li, C., Xie, C., Zhang, B., Chen, C., & Han, J. (2018). Deep Fisher discriminant learning for mobile hand gesture recognition. *Pattern Recognition*, *77*, 276-288

Recurrent Neural Networks (RNN)

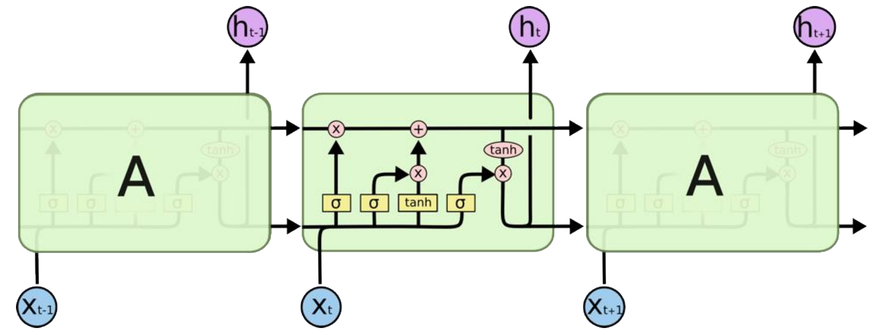


RNNs have loops → An unrolled RNNS.

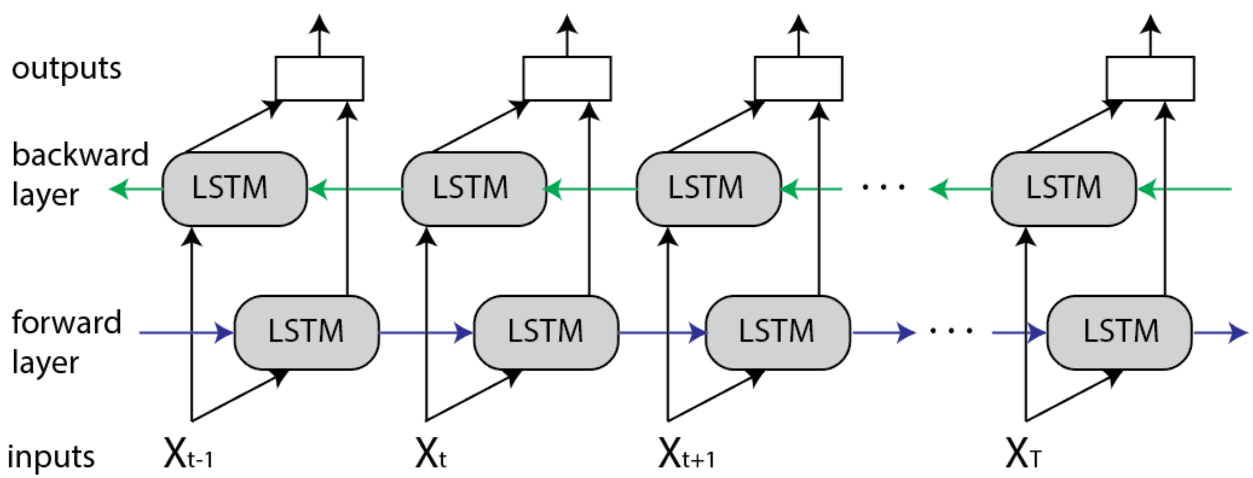
Long Short Term Memory Networks (LSTM)



Usual RNNs



LSTM



Bidirectional LSTM

Multilayer RNN

other architectures

Vanilla RNN:

$$h_t^l = \tanh W^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix}$$

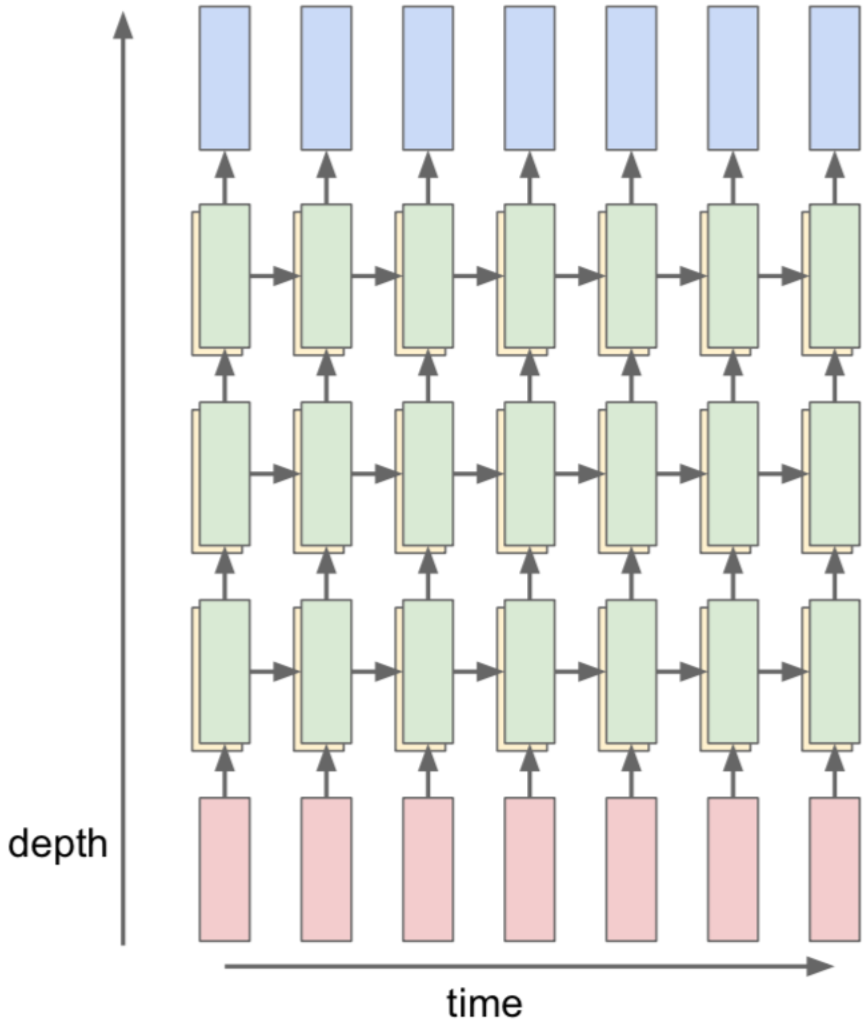
$h \in \mathbb{R}^n$ $\quad W^l \ [n \times 2n]$

LSTM:

$W^l \ [4n \times 2n]$

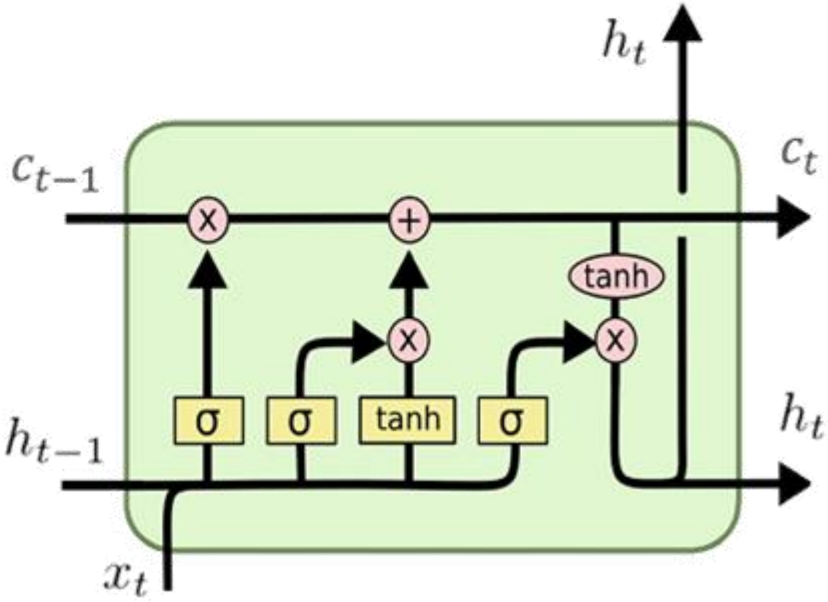
$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \text{sigm} \\ \text{sigm} \\ \text{sigm} \\ \tanh \end{pmatrix} W^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix}$$

$$c_t^l = f \odot c_{t-1}^l + i \odot g$$
$$h_t^l = o \odot \tanh(c_t^l)$$



LSTM

other architectures

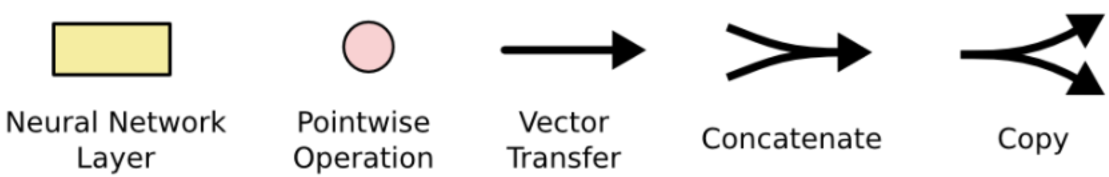


LSTM
(Long-Short Term Memory)

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \text{sigm} \\ \text{sigm} \\ \text{sigm} \\ \tanh \end{pmatrix} W^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix}$$

$$c_t^l = f \odot c_{t-1}^l + i \odot g$$

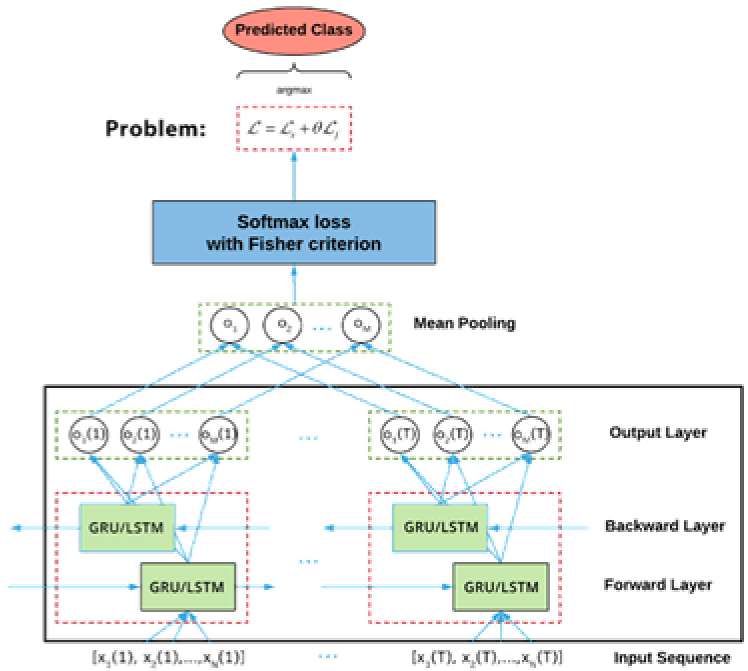
$$h_t^l = o \odot \tanh(c_t^l)$$



* See **additional materials** on https://elearning.unimib.it/course/view.php?id=37919

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

Bidirectional Recurrent Neural Networks
F-BLSTM (Fisher criterion Bidirectional Long-Short Term Memory)

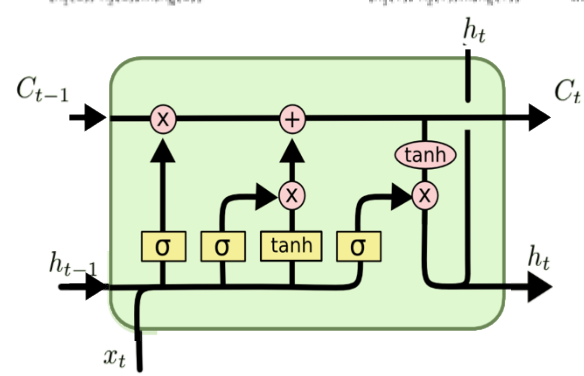


LSTMs: the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way

We have M=64 hidden layers that are fully connected to the backwards and forwards LSTM units. N=6 and T=500.

forget gate layer

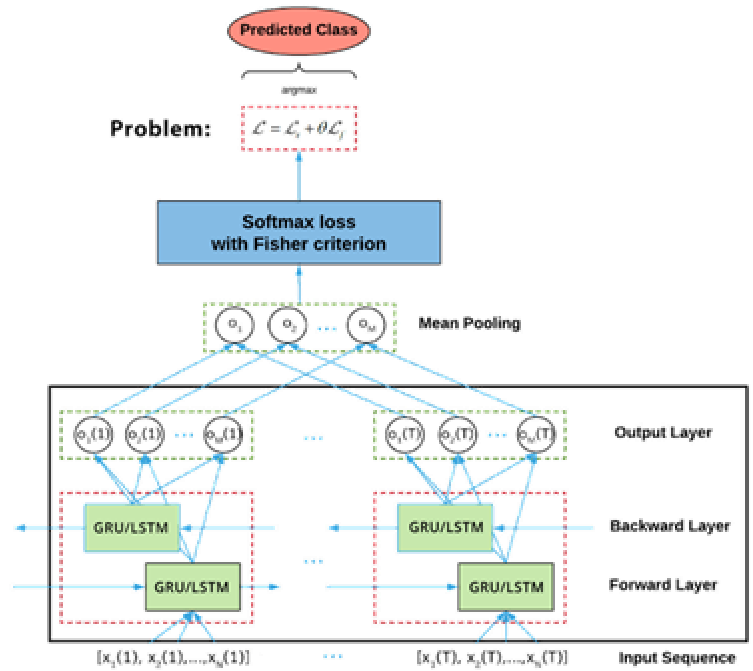
$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$



Bidirectional Recurrent Neural Networks

F-BLSTM (Fisher criterion Bidirectional Long-Short Term Memory)

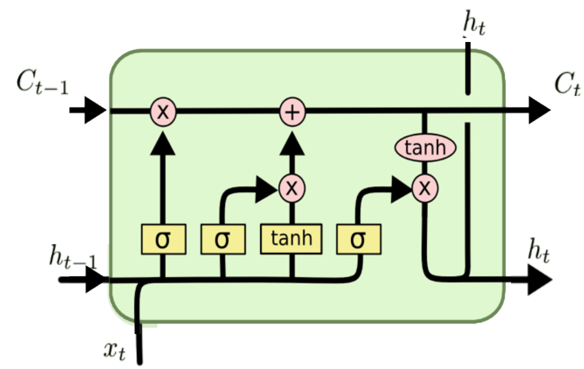
LSTMs: the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way



input gate and tanh layers

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$

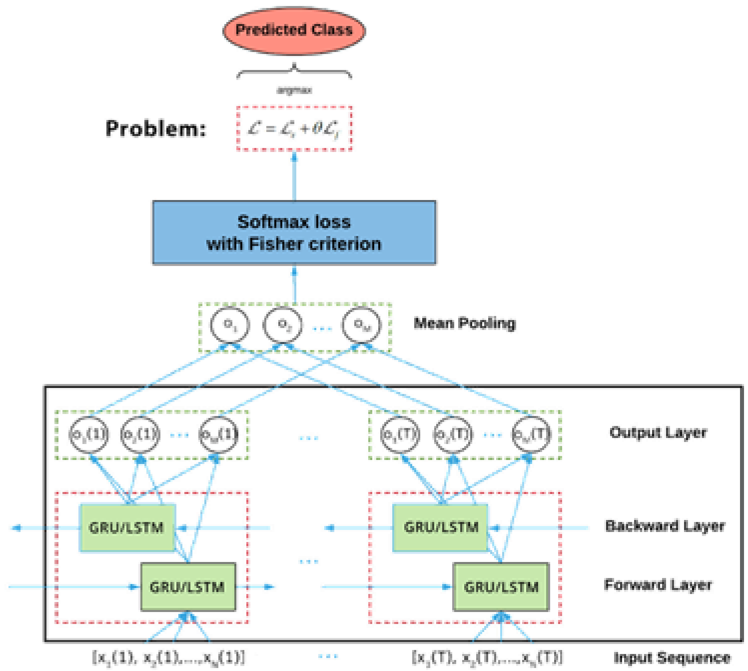
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$



Bidirectional Recurrent Neural Networks

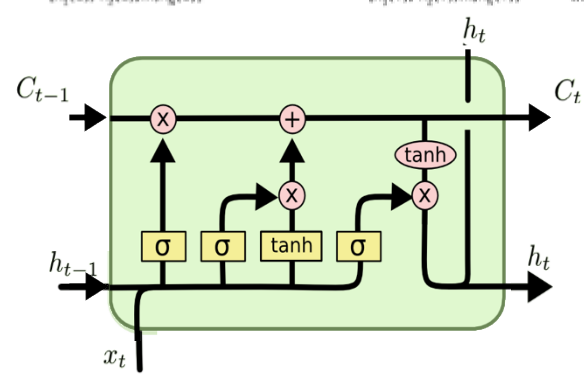
F-BLSTM (Fisher criterion Bidirectional Long-Short Term Memory)

LSTMs: the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way



Updating the cell state

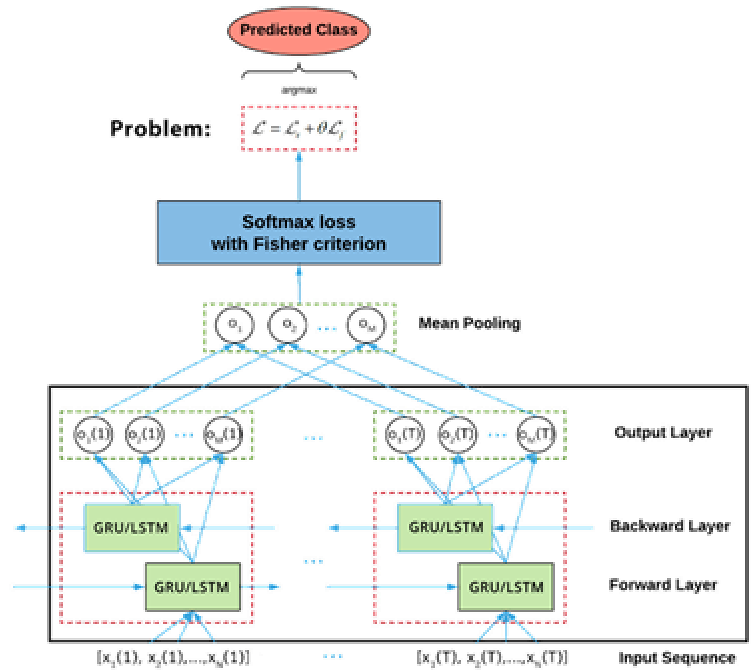
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$



Bidirectional Recurrent Neural Networks

F-BLSTM (Fisher criterion Bidirectional Long-Short Term Memory)

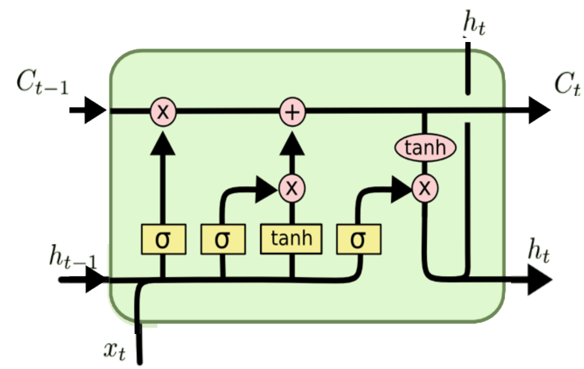
LSTMs: the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way



Output layer

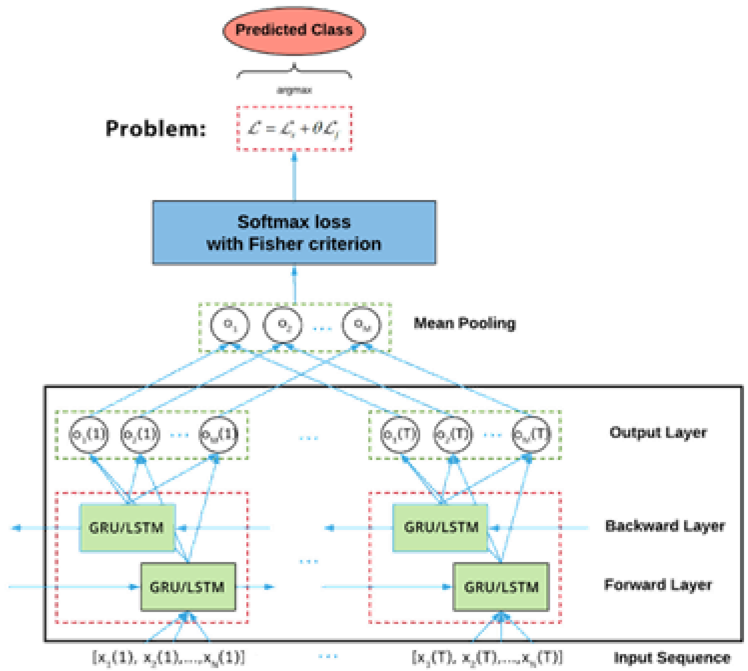
$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$



Bidirectional Recurrent Neural Networks

F-BGRU (Fisher criterion Bidirectional Gated Recurrent Unit)



GRU was proposed to make each recurrent unit to adaptively capture dependence of different time scales

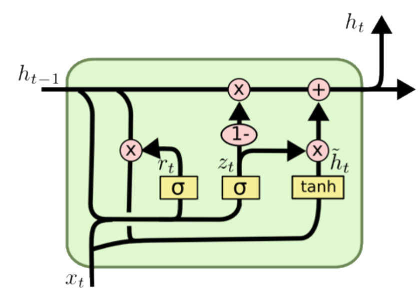
It combines the forget and input gates into a single "update gate."

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$



Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

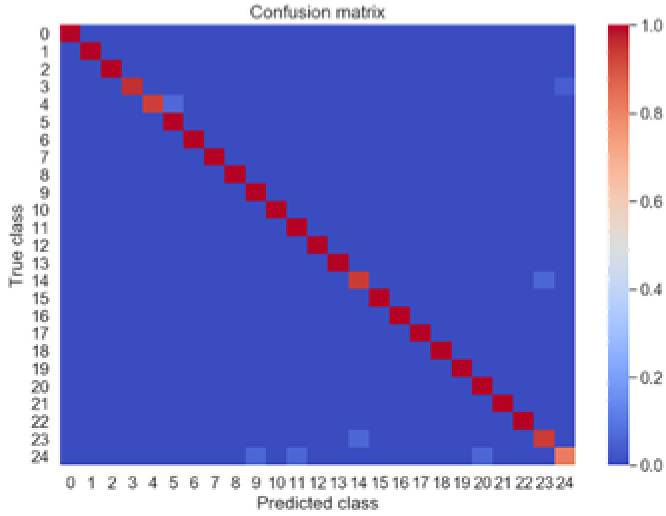
Arm gesture recognition results

Concerning the data split, for the gesture recognition experiments a fivefold cross validation and a leave one subject out cross validation (**LOSO**) are performed

	Accuracy		F1-measure	
	Fivefold CV	LOSO	Fivefold CV	LOSO
BLSTM	95.75 ($\pm$ 0.6)	94.35 ($\pm$ 5.1)	95.83 ($\pm$ 0.6)	94.80 ($\pm$ 4.7)
BGRU	96.24 ($\pm$ 0.5)	95.37 ($\pm$ 4.6)	96.29 ($\pm$ 0.5)	95.63 ($\pm$ 4.8)
F-BLSTM	95.90 ($\pm$ 0.8)	95.00 ($\pm$ 4.5)	96.00 ($\pm$ 0.8)	95.38 ($\pm$ 4.3)
F-BGRU	**96.63** ($\pm$ 0.3)	**95.58** ($\pm$ 4.2)	**96.70** ($\pm$ 0.3)	**95.67** ($\pm$ 4.3)

It is possible to see that the use of the Fisher criterion is able to improve the performance

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$



Normalized confusion matrix of the F-BGRU model trained for gesture recognition.

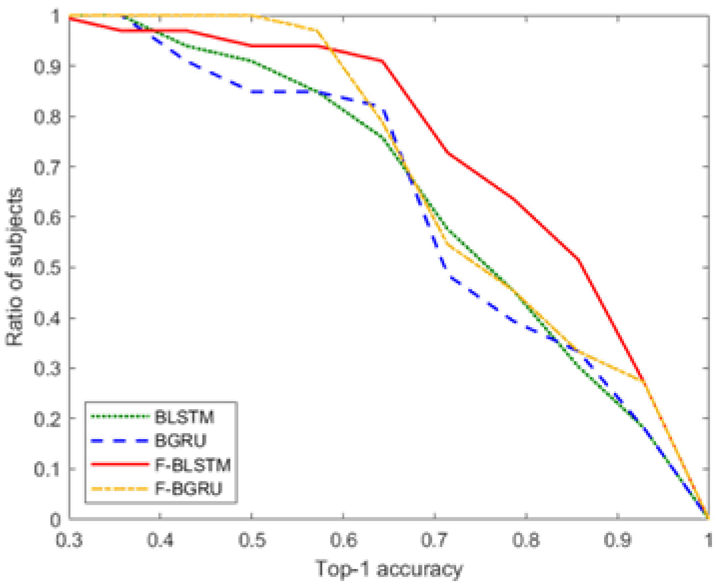
User identification results

Concerning the data split, for the user identification experiment a fivefold cross validation is performed

	Top-1 acc.	Top-3 acc.	Top-5 acc.	F1-measure
BLSTM	82.61 (± 2.9)	93.48 (± 1.4)	96.32 (± 0.9)	83.00 (± 2.7)
BGRU	80.65 (± 2.2)	91.92 (± 1.2)	94.76 (± 0.7)	81.00 (± 2.1)
F-BLSTM	**85.24** (± 1.3)	**94.35** (± 0.4)	**96.68** (± 0.2)	**85.57** (± 1.3)
F-BGRU	83.00 (± 0.8)	93.23 (± 0.4)	96.00 (± 0.3)	83.42 (± 0.7)

Similarly to the first experiment the use of the Fisher criterion improves the results for both BLSTM and BGRU

Curves representing the ratio of subjects having a top-1 accuracy above a given threshold in the range [0.3;1] for all the methods



User authentication

Concerning the data split, for the user authentication experiment a different split is used by considering the data of 28 of the users as the training set and the data of the remaining five users as the test set.

For the user authentication step the following distance metrics are used, both with and without l2 normalization.

1) Euclidean distance (Euc).
2) Normalized Euclidean distance (N-Euc).
3) Standardized Euclidean distance (S-euc).

User authentication process

User authentication

- Once the F-BLSTM model has been trained for user identification, the softmax activation is discarded, and feature extraction using the unnormalized log probabilities is performed, thus producing a 128-dimensional representation for each gesture.
- After the features have been extracted for all the training and test datasets the pairwise distances are computed using the three distance metrics described in the previously.
- For each metric considered, both with and without l2 normalization to unit vectors, user authentication performance are measured by computing the EER.

Equal Error Rate

$$P_{\text{Miss}}(\theta_{\text{EER}}) = P_{\text{FA}}(\theta_{\text{EER}})$$

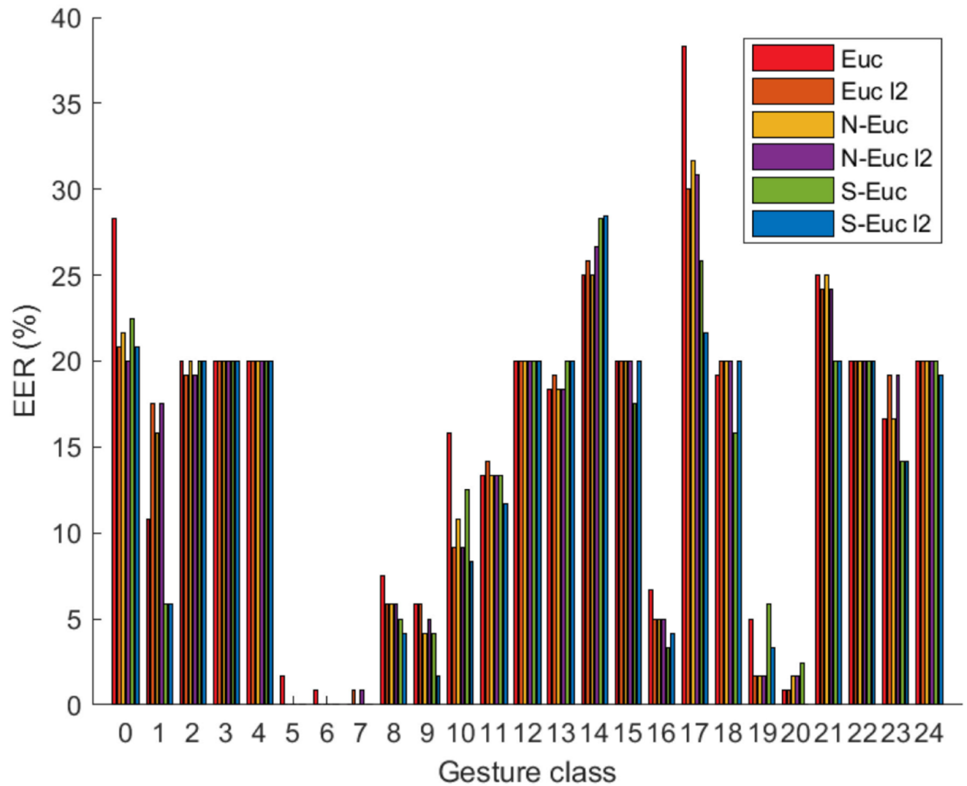
$$\theta = \text{decision threshold}$$

User authentication results

	Gesture independent	Gesture dependent
Euc	38.38 ($\pm$ 0.77)	15.17 ($\pm$ 0.74)
Euc-l_2	<u>36.24</u> ($\pm$ 0.87)	14.37 ($\pm$ 0.45)
N-Euc	36.62 ($\pm$ 0.89)	14.27 ($\pm$ 0.74)
N-Euc-l_2	<u>36.25</u> ($\pm$ 0.85)	14.33 ($\pm$ 0.45)
S-Euc	37.69 ($\pm$ 0.55)	13.46 ($\pm$ 0.53)
S-Euc-l_2	**35.62** ($\pm$ 0.78)	**12.94** ($\pm$ 0.29)

The standardized Euclidean distance with l2 normalization (S-Euc-l2) is the variant of Euclidean distance that achieves the best results in both the gesture independent and gesture independent setups.

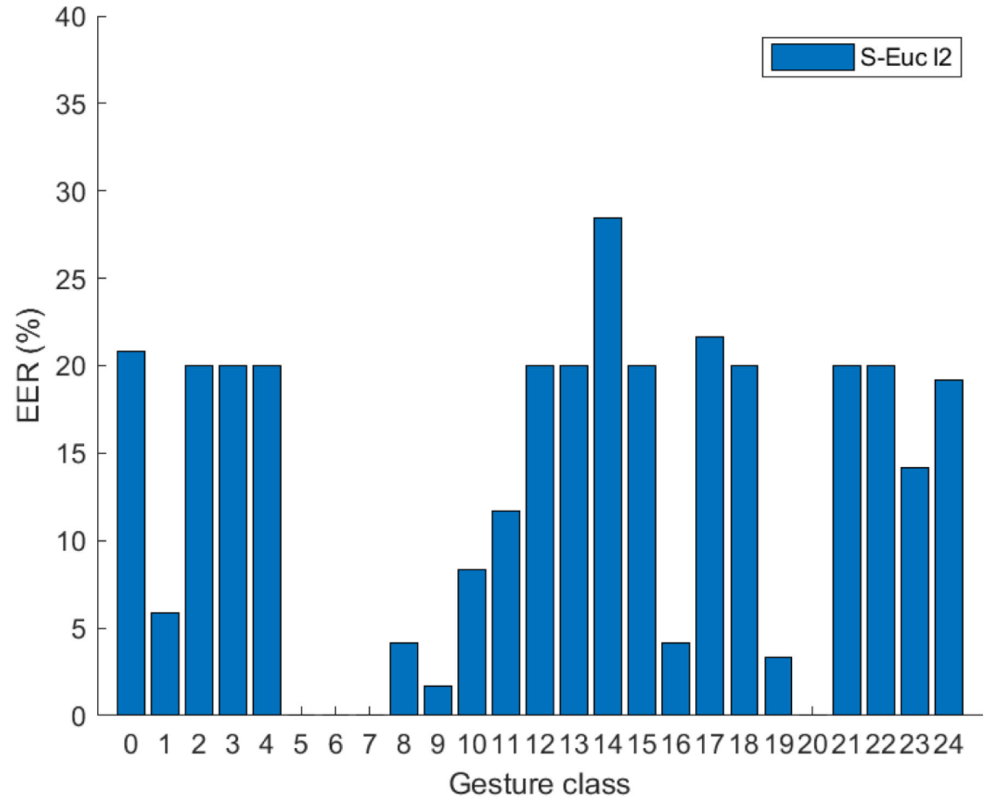
User authentication results



Considering the standardized Euclidean distance with l2 normalization (i.e., S-Euc-l2), the user authentication performs noticeably better for some classes of gestures: the recognition of users performing gestures belonging to class 5, 6, 7, and 20 has a zero EER; while five other classes of gestures (i.e., class 1, 8, 9, 16, and 19) have a lower EER compared to the other classes.

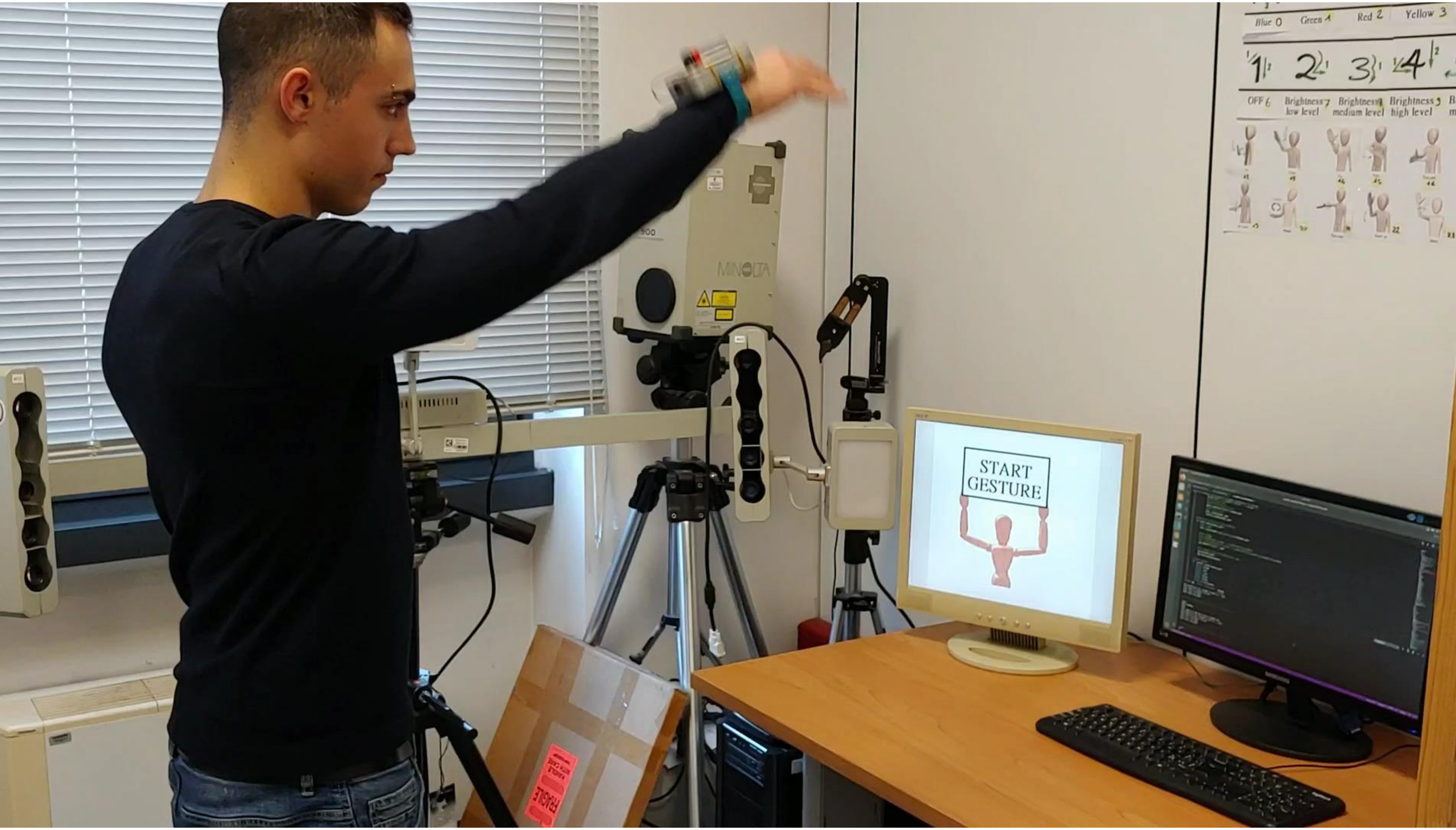
Gesture authentication results

	Gesture dependent robustness
Euc	19.60 (± 0.45)
Euc l2	18.66 (± 0.12)
N-Euc	18.96 (± 0.05)
N-Euc l2	18.85 (± 0.11)
S-Euc	18.43 (± 0.29)
S-Euc l2	**17.37** (± 0.35)

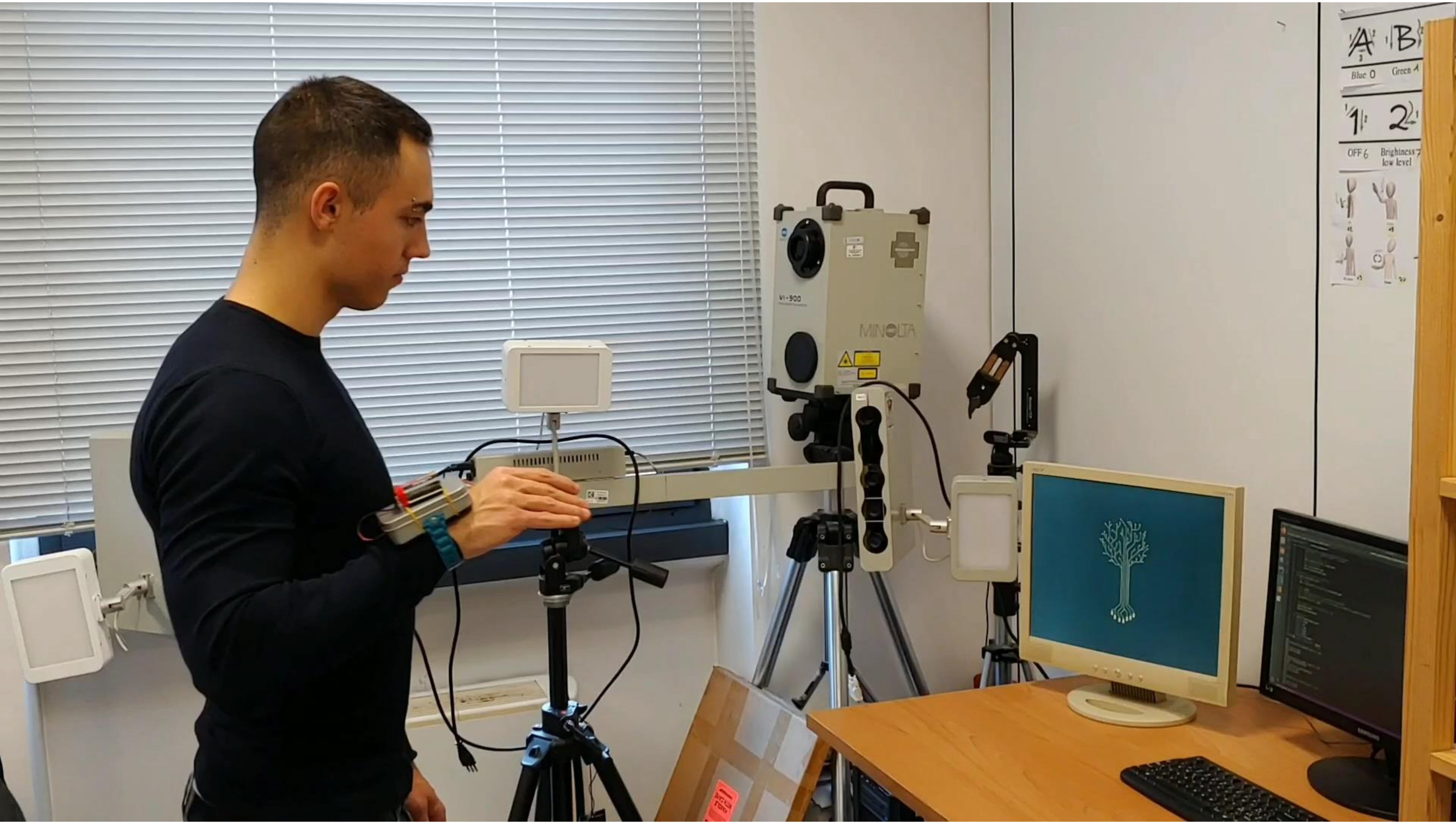


gesture authentication: we define a correct authentication when for a gesture performed by a user, the closest gesture belongs to the same class independently from the user that performed it.

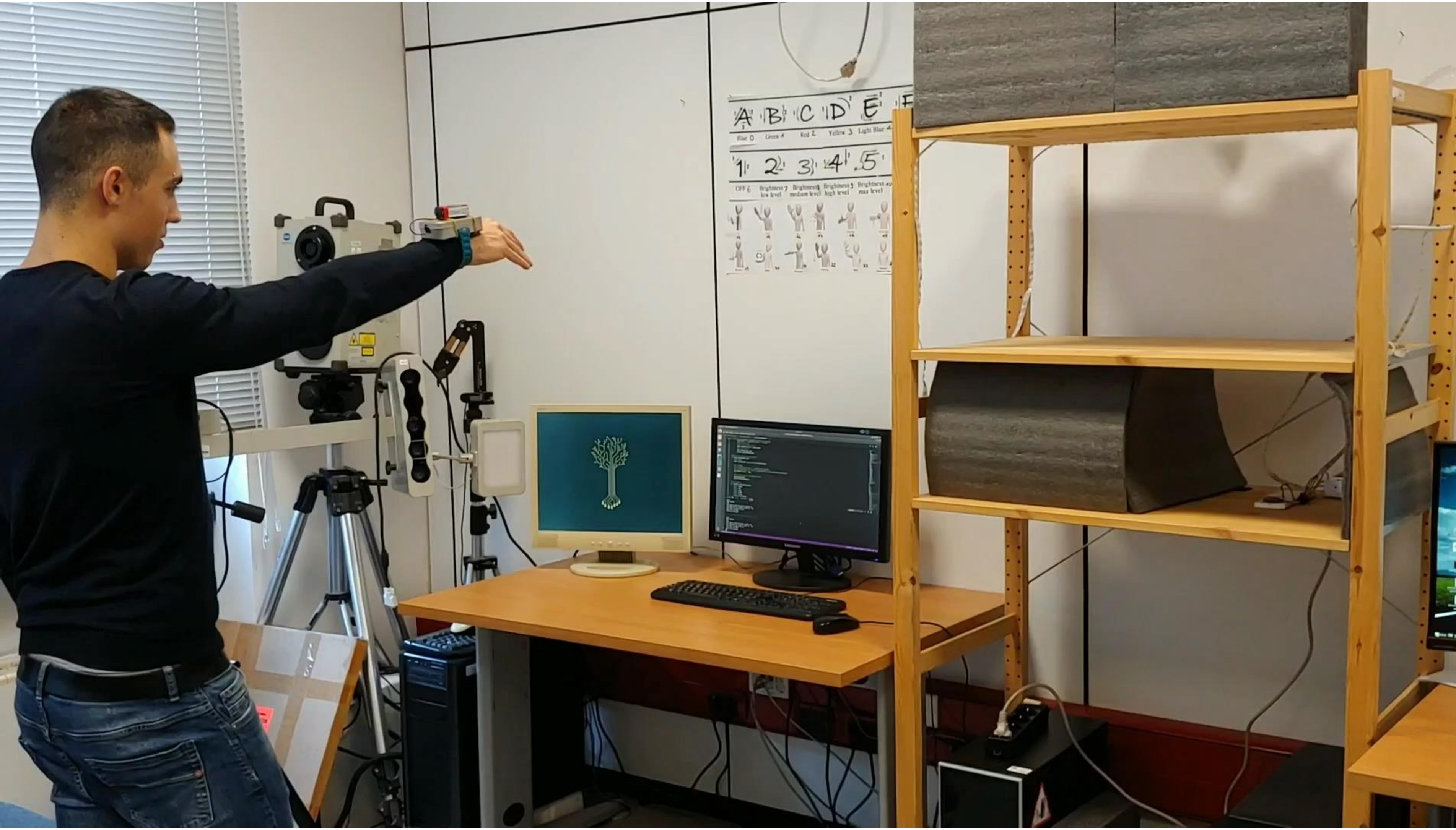
Prototype – demo 4



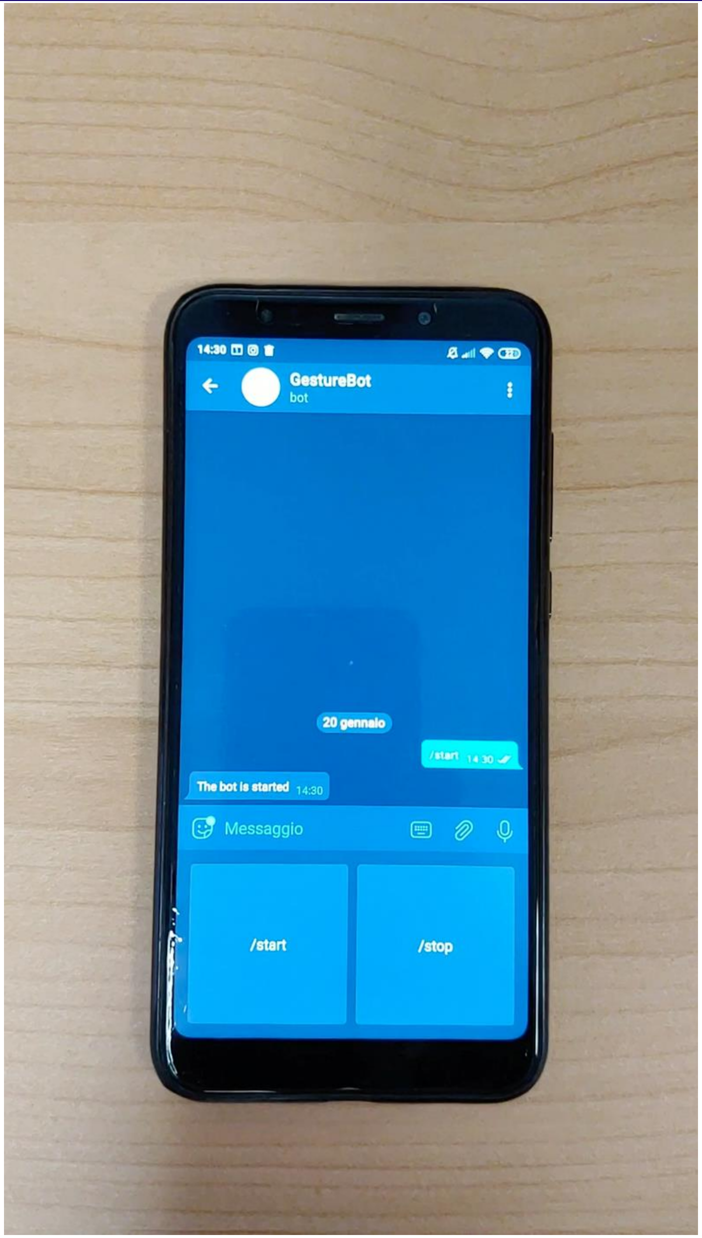
Prototype – demo 1



Prototype – demo 3



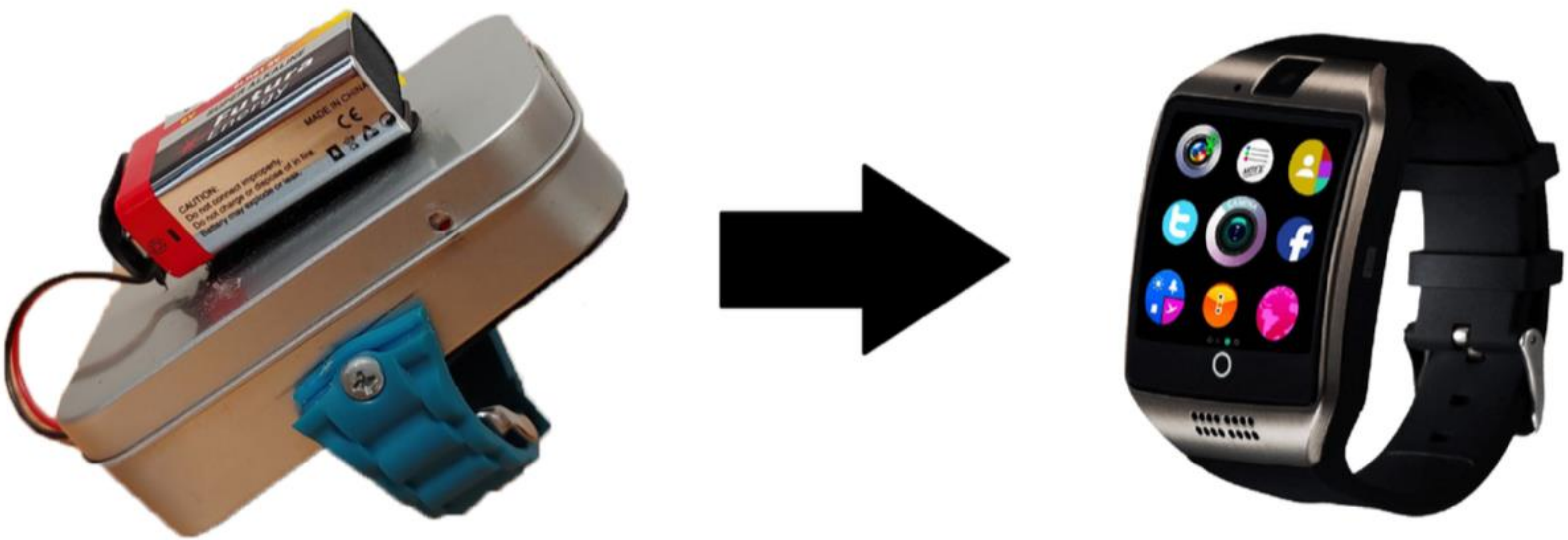
Prototype - demo 2



Biometric recognition

Using the bracelet to unlock private devices

Privacy preserved!!

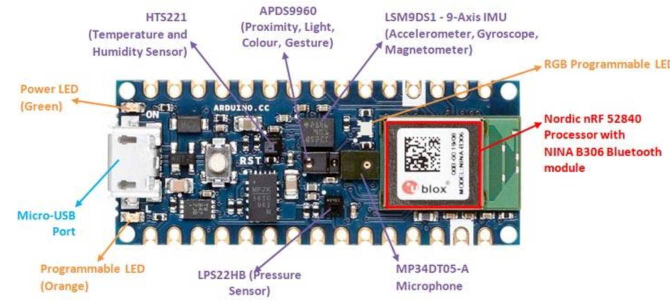


Biometric recognition

Using the bracelet to unlock private devices

Privacy preserved!!

TinyML



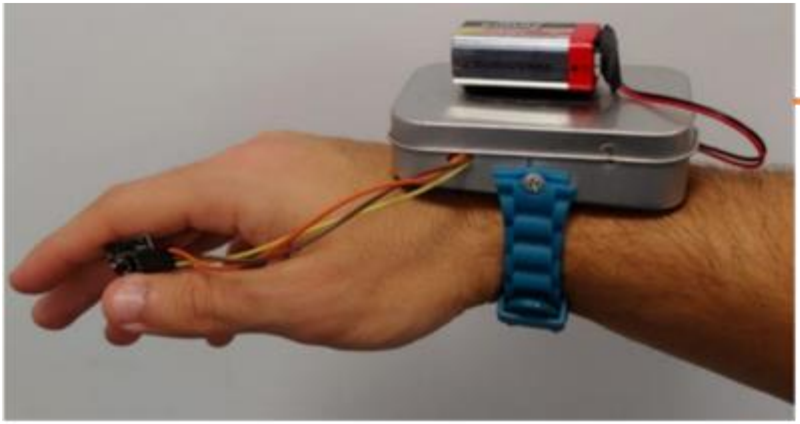
Biometric recognition

The Arduino Nano 33 BLE is based on the nRF52840 microcontroller.

Microcontroller	nRF52840 (datasheet)
Operating Voltage	3.3V
Input Voltage (limit)	21V
DC Current per I/O Pin	15 mA
Clock Speed	64MHz
CPU Flash Memory	1MB (nRF52840)
SRAM	256KB (nRF52840)
EEPROM	none
Digital Input / Output Pins	14
PWM Pins	all digital pins
UART	1
SPI	1
I2C	1
Analog Input Pins	8 (ADC 12 bit 200 ksamples)
Analog Output Pins	Only through PWM (no DAC)
External Interrupts	all digital pins
LED_BUILTIN	13
USB	Native in the nRF52840 Processor
Length	45 mm
Width	18 mm
Weight	5 gr (with headers)

NodeMCU Technical Sp	
	Official NodeMCU Amica
Microcontroller	ESP-8266 32-bit
NodeMCU Model	Amica
NodeMCU Size	49mm x 26mm
Pin Spacing	**0.9" (22.86mm)**
Clock Speed	80 MHz
USB to Serial	CP2102
USB Connector	Micro USB
Operating Voltage	3.3V
Input Voltage	4.5V-10V
Flash Memory/SRAM	4 MB / 64 KB
Digital I/O Pins	11
Analog In Pins	1
ADC Range	0-3.3V
UART/SPI/I2C	1 / 1 / 1
WiFi Built-In	802.11 b/g/n
Temperature Range	-40C - 125C

Wristband for arm gesture recognition



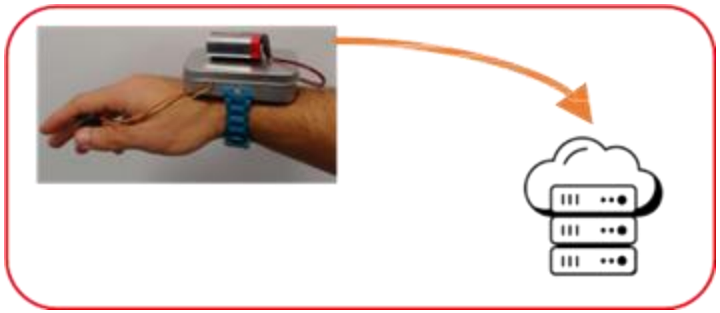
Wristband equipped with an Inertial
Measurement Unit (IMU)

MQTT

Server for performing **deep model inference**
for gesture recognition

S. Bianco, P. Napoletano, A. Raimondi, and M. Rima, "U-wear: User recognition on wearable devices through arm gesture," *IEEE Transactions on Human-machine Systems*, vol. 52, no. 4, pp. 713–724, 2022.

Using smartwatches



Google Pixel Watch 2

Qualcomm 5100 chip coupled with a Cortex M33 coprocessor

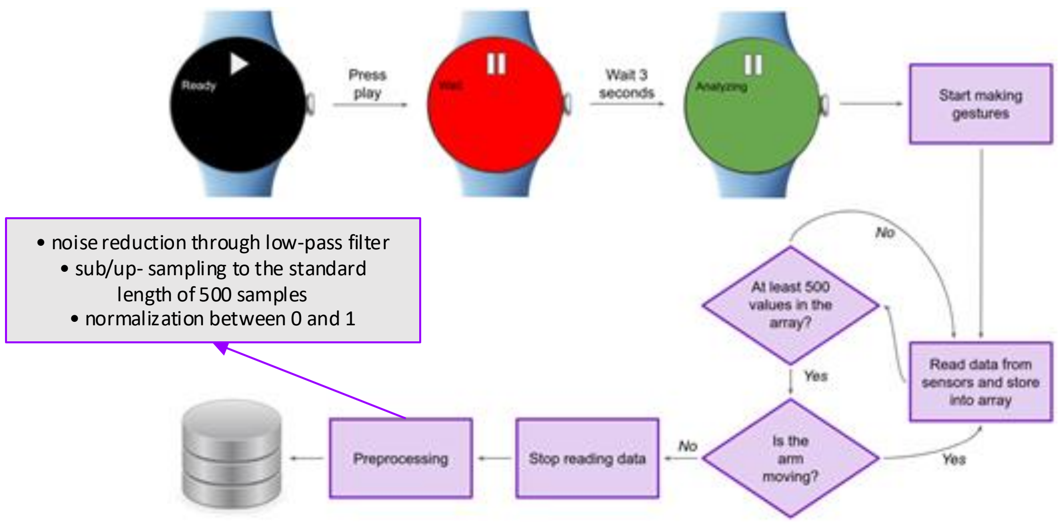
2GB SDRAM and 32GB of eMMC flash memory

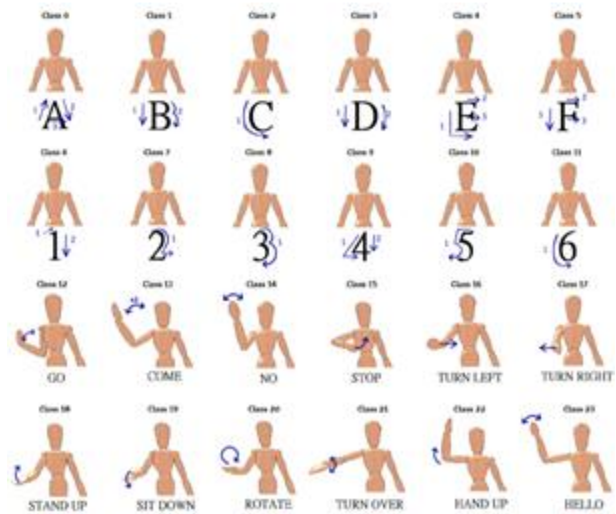
<u>**Challenges**</u>

A direct porting is actually **not feasible** due to the difference in the device’s IMU

The model requires optimization and size reduction to ensure compatibility and efficient execution on the resource-constrained device

Dataset





- Dataset of 875 arm gestures
- 25 classes (24 gestures + reject class)
- 7 subjects
- 5 repetitions

S. Bianco, P. Napoletano, A. Raimondi, and M. Rima, "U-wear: User recognition on wearable devices through arm gesture," *IEEE Transactions on Human-machine Systems*, vol. 52, no. 4, pp. 713–724, 2022.

Post-Training Quantization results

PTQ is a technique used to convert pretrained models to **a more compact format with only minimal impact on model accuracy**

Dynamic Range PTQ (DR-PTQ)

Full Integer PTQ (FI-PTQ)

- Integer with Float Fallback (IFF FI-PTQ)
- Integer only (IO FI-PTQ)

Float16 PTQ (FP16-PTQ)

Int8 weights and Int16 activations PTQ (16x8-PTQ)

TensorFlow Lite

The results are estimated for one of the checkpoints that performed best on the test, namely, the epoch 88 checkpoint of the 5-fold CV experiment

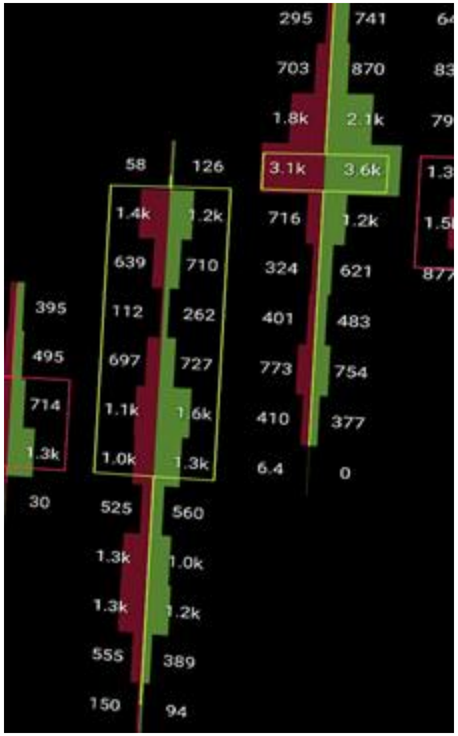
Conversion type	Weight Bit-width	Activation Bit-width	Weight (MB)	Acc. (%)	Prec.	Rec.	F1
TF	32	32	35.1	89.14	0.91	0.89	0.90
TFLite	32	32	2.7	89.14	0.91	0.89	0.90
IO FI-PTQ	8	8	14.7	86.84	0.88	0.87	0.88
IFF FI-PTQ	8	8	14.7	87.34	0.89	0.88	0.88
16x8-PTQ	8	16	3.9	88.57	0.90	0.89	0.89
FP16-PTQ	16	32	2.5	89.14	0.91	0.89	0.90
DR-PTQ	8	8	2.4	89.71	0.91	0.90	0.90

Application profiling

Profiling revealed that the application imposed a **low CPU load**
- Booting and data recording involved approximately 3% CPU usage
- The inference process caused a higher CPU load of about 20%

The **base memory usage** was recorded at 37 MB
- + 1 MB during the data collection phase
- + 5 MB during data preprocessing and inference

The **baseline energy consumption** is around 100mA
- A peak of 280mA is recorded at the start of gesture recording (about 0.5 seconds)
- The consumption then stabilizes at around 120mA
- For data preprocessing and inference, a consumption spike of approximately 350mA is registered



QUESTIONS?