



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN INGEGNERIA INFORMATICA

DATASET APERTI PER IL RICONOSCIMENTO DI ATTIVITÀ UMANE

Relatore:

Carlo Fantozzi

Laureando:

Fabio Zanini

ANNO ACCADEMICO: 2021/2022

Data di laurea: 21 Settembre 2022

Sommario

Il riconoscimento di attività umane (HAR), sempre più popolare all'interno della comunità scientifica, si basa su tecniche di machine learning e di deep learning. Queste tecniche necessitano di grandi quantità di dati per l'addestramento degli algoritmi che generano. Queste grandi quantità di dati si trovano in dataset, raccolte di dati che possono essere aperte o meno, e che possono avere caratteristiche molto diverse l'una dall'altra.

Le tre principali tipologie di dataset aperti per il riconoscimento di attività umane sono: dataset contenenti dati da sensori, dataset contenenti segnali radio (RF) e segnali WiFi, dataset contenenti video.

Della prima tipologia fanno parte i dataset WISDM, Berkeley MHAD, PAMAP2, Opportunity e DSADS. Alla seconda appartengono OPERAnet, Mendeley e WiAR. Infine, della terza categoria fanno parte UCF101, UCF50, UCF11, UCF Sports Action e HMDB51. Ognuno di questi dataset contiene un numero diverso di azioni svolte dai soggetti, un diverso quantitativo di dati e dimensioni differenti.

Ogni dataset è presentato, da chi lo ha creato, in un articolo scientifico. Per confrontare i dataset, è stato confrontato il numero di citazioni, dato preso da Google Scholar, di ognuno di questi articoli e, inoltre, sono stati confrontati numero di citazioni e tipologia dei tre articoli più citati che utilizzano ognuno dei dataset.

Altri confronti svolti hanno analizzato il numero di azioni e di soggetti presenti nel dataset e il numero di elementi del dataset.

Così facendo, è stato possibile avere una visione d'insieme del materiale e della letteratura disponibile nell'ambito dei dataset aperti per il riconoscimento di attività umane.

Indice

1	Introduzione	1
1.1	Reperimento delle informazioni	1
2	Tipologie di dataset selezionati	2
3	Dataset contenenti dati da sensori	3
3.1	WISDM	3
3.2	Berkeley MHAD	4
3.3	PAMAP2	6
3.4	Opportunity	7
3.5	DSADS	8
4	Dataset con segnali radio e WiFi	10
4.1	OPERAnet	10
4.2	Mendeley	10
4.3	WiAR	11
5	Dataset contenenti video	12
5.1	YouTube Action (UCF11)	12
5.2	UCF50	13
5.3	UCF101	14
5.4	UCF Sports Action	16
5.5	HMDB51	17
6	Confronti tra dataset	19
6.1	Numero di elementi	19
6.2	Numero di azioni e di soggetti	21
6.3	Numero di citazioni	22
7	Conclusioni	27

1 Introduzione

Il riconoscimento di attività umane (Human Activity Recognition, HAR) sta diventando sempre più popolare. Tale interesse è dovuto al fatto che questo tipo di studio può essere utilizzato in un insieme molto vario di ambiti: dalle attività quotidiane più comuni, come la pratica di attività sportive, al monitoraggio della salute personale.

Alla base del riconoscimento si trova l'utilizzo di tecniche di machine learning e deep learning, che riescono a formulare algoritmi in grado, appunto, di riconoscere attività umane in condizioni anche molto variabili tra caso e caso. Queste tecniche necessitano di grandi quantitativi di dati, i quali devono rappresentare varie tipologie di attività, in varie condizioni ambientali, svolte da soggetti diversi, con caratteristiche fisiche diverse, e così via, in modo da ampliare lo spettro di azione dell'algoritmo che ne deriverà.

Alcune delle sfide che gli algoritmi per il riconoscimento di attività umane devono affrontare sono: la varietà e la complessità delle attività quotidiane; la variabilità nello svolgimento dell'attività da parte di uno stesso soggetto o di soggetti diversi; il compromesso, più che mai attuale, tra prestazioni e privacy; l'efficienza dell'algoritmo, per essere utilizzato all'interno di dispositivi portatili ed embedded; la difficoltà nell'annotare i dati raccolti [1].

Come già accennato, per l'addestramento e il testing degli algoritmi di HAR sono necessarie grandi quantità di dati. Tali dati possono essere raccolti principalmente attraverso: sensori ambientali, sensori integrati, fotocamere (e quindi video) e segnali radio o WiFi.

Parte integrante del lavoro, quindi, sta nella ricerca e nell'utilizzo di dataset di grandi dimensioni, che siano affidabili, che possano soddisfare le necessità degli sviluppatori e che, possibilmente, siano aperti, cioè pubblicamente disponibili. L'importanza di questi dataset risiede nel fatto che i ricercatori non devono ripetutamente disperdere energie nella raccolta dati, e possono confrontare le prestazioni di algoritmi diversi sugli stessi dati. Da cui l'obiettivo di questa tesina: elencare e analizzare i principali dataset aperti per il riconoscimento di attività umane, concentrandosi su quelli di grandi dimensioni e che siano citati in un grande numero di articoli scientifici, mettendo in evidenza le differenze tra di essi, in modo da fornire una panoramica del materiale disponibile.

1.1 Reperimento delle informazioni

Per selezionare i dataset che verranno analizzati nel seguito, si è partiti da alcune tabelle ed elenchi presenti negli articoli scientifici [1], [2] e [3]. Per ogni dataset è stato controllato, con l'utilizzo di Google Scholar, il numero di articoli scientifici in cui esso è stato citato e l'importanza di tali articoli (importanza attribuita in base al numero di citazioni dell'articolo). In questo modo è stata effettuata una selezione dei soli dataset ampiamente citati, ed è stato poi possibile procedere con l'analisi delle caratteristiche di ognuno e il reperimento delle informazioni necessarie ad effettuare i confronti tra di essi. Tali informazioni sono state reperite all'interno dei siti web dei creatori dei dataset e tramite gli articoli che li citano o che li descrivono direttamente.

Dopo aver analizzato ciascuno dei dataset selezionati, è stato quindi possibile confrontare i dataset secondo vari criteri, quali tipologie di dati contenuti o dimensioni.

2 Tipologie di dataset selezionati

I dataset selezionati per la ricerca possono essere raggruppati in tre gruppi distinti:

- dataset contenenti dati da sensori,
- dataset contenenti segnali radio (RF) e segnali WiFi,
- dataset contenenti video.

Dataset contenenti dati da sensori I dataset di questa prima tipologia sono composti da dati derivanti da varie tipologie di sensori: accelerometri, giroscopi e magnetometri, principalmente, ma anche GPS, sensori di luminosità, di temperatura, e così via. Questi possono essere direttamente posizionati sui soggetti che svolgono l'azione da registrare, oppure possono essere integrati in uno smartphone, il quale viene fatto indossare al soggetto.

La possibilità di sfruttare gli smartphone permette di semplificare il processo di raccolta di dati. Inoltre, rende possibile personalizzare gli algoritmi di riconoscimento e il risultato derivante in base alle abitudini dell'utente, il quale indosserà quasi costantemente uno smartphone, e quindi dei sensori costantemente utilizzabili. Questa personalizzazione è utile, ad esempio, per creare dei profili personalizzati in base alla quantità di esercizio fisico svolta dall'utente e alle sue abitudini giornaliere, in modo da monitorarne lo stato di salute fisica [4].

Dataset con segnali radio e WiFi I dataset in questione contengono misure di segnali radio, ottenute attraverso l'utilizzo di dispositivi WiFi di vario tipo, come per esempio WiFi Channel State Information (CSI), Passive WiFi Radar (PWR), Ultra-wideband (UWB) [5].

Il metodo più comune è quello basato sulle informazioni di canale di un segnale WiFi (CSI), ossia l'analisi delle proprietà del canale di comunicazione tra ricevitore e trasmettitore, per evidenziare eventuali interferenze, dovute, nel nostro caso, all'interazione del segnale con il soggetto osservato.

In alcuni casi, i segnali WiFi vengono affiancati a segnali di altro tipo, per esempio derivanti da un Kinect [5].

Tuttavia, non sono ancora presenti dataset aperti di grandi dimensioni e molto citati che contengano questo tipo di dati. Per questo, nell'analisi che verrà svolta, questo tipo di dataset assumerà minore importanza rispetto alle altre due tipologie.

Dataset contenenti video I set di questo tipo contengono dati sotto forma di video. I video possono essere reperiti su piattaforme online, quali YouTube, oppure possono essere registrati da chi svolge la ricerca.

Le sfide principali nella creazione e nell'utilizzo di tali dataset sono due: il numero di classi di azioni registrate è spesso basso rispetto alla ricchezza e al numero di azioni naturali che l'uomo compie nel corso della sua vita; spesso i video sono registrati in condizioni artificiali e non reali [6].

Analizzando i frame e le relazioni tra i frame presenti nei video, è possibile sviluppare algoritmi che siano in grado di riconoscere le azioni per cui sono stati progettati.

3 Dataset contenenti dati da sensori

3.1 WISDM

WISDM (Wireless Sensor Data Mining) [4] è un dataset creato nel 2010 dal WISDM Lab, un laboratorio di ricerca che si occupa di raccogliere dati dai sensori di smartphone o altri dispositivi mobili e di studiarli per ottenere informazioni utili. Tale progetto, guidato dal Dr. Gary Weiss, membro accademico della facoltà di Computer and Information Science alla Fordham University, conta più di una dozzina di membri, la maggior parte dei quali sono studenti della Fordham University.

Il dataset, scaricabile al link [7], contiene dati raccolti da 51 soggetti, ad ognuno dei quali è stato chiesto di svolgere 18 azioni per 3 minuti ognuna [8]. Ogni soggetto ha indossato uno smartwatch (LG G Watch con Android Wear 1.5) sul polso del braccio dominante, e uno smartphone (Google Nexus 5/5x o Samsung Galaxy S5) nella tasca dei pantaloni.

I dati raccolti sono quelli dell'accelerometro (in m/s^2) e del giroscopio (in radianti/s) presenti sia nello smartphone che nello smartwatch. Tali dati sono stati raccolti ad una frequenza di 20 Hz.

Codice	Attività
A	Camminare
B	Fare jogging
C	Salire/scendere le scale
D	Stare seduti
E	Stare in piedi
F	Scrivere su una tastiera
G	Lavarsi i denti
H	Mangiare una zuppa
I	Mangiare delle patatine
J	Mangiare della pasta
K	Bere da una tazza
L	Mangiare un panino
M	Calcciare un pallone da calcio
O	Prendere una pallina da tennis
P	Dribbling con un pallone da basket
Q	Scrivere
R	Applaudire
S	Piegare vestiti

Tabella 1: Le 18 azioni di WISDM

Nei file, divisi per sensore e per attività, ogni riga è una lettura del sensore, ed è così formata: *ID del soggetto* , *Codice dell'attività svolta* , *Timestamp*, *x*, *y*, *z*; . Il numero totale di righe, e quindi di letture, per ogni sensore, è il seguente.

- Accelerometro smartphone: 4,804,403.
- Giroscopio smartphone: 3,608,63.
- Accelerometro smartwatch: 3,777,046.

- Giroscopio smartwatch: 3,440,342.

Il dataset, oltre ai dati grezzi, contiene anche i dati etichettati.

3.2 Berkeley MHAD

Il dataset Berkeley MHAD (Multimodal Human Action Database) [9] è frutto della collaborazione tra il Teleimmersion Lab della University of California a Berkeley e il Center for Imaging Science della Johns Hopkins University. Esso è stato pubblicato nel 2013 ed è scaricabile dal sito ufficiale [10]. Questo dataset contiene 11 azioni svolte da 7 soggetti maschi e 5 femmine, con età compresa tra i 23 e i 30 anni, ad eccezione di un soggetto più anziano. Ogni soggetto ha effettuato 5 ripetizioni per ogni azione. Inoltre, ogni soggetto è stato fatto posare fermo con le braccia aperte (a T), in modo da poterne ricavare la lettura della forma del corpo [10].

Le azioni registrate comprendono: azioni con movimenti sia nella parte superiore che inferiore del corpo, azioni con grande dinamismo nella parte superiore e azioni con grande dinamismo nella parte inferiore.

Per l'acquisizione dei dati è stato utilizzato un sistema di acquisizione multi-modale. In particolare, ogni azione è stata catturata simultaneamente da: un sistema ottico di cattura del movimento, 4 serie di videocamere per la visione stereo, due Kinect Microsoft, 6 accelerometri wireless e 4 microfoni.

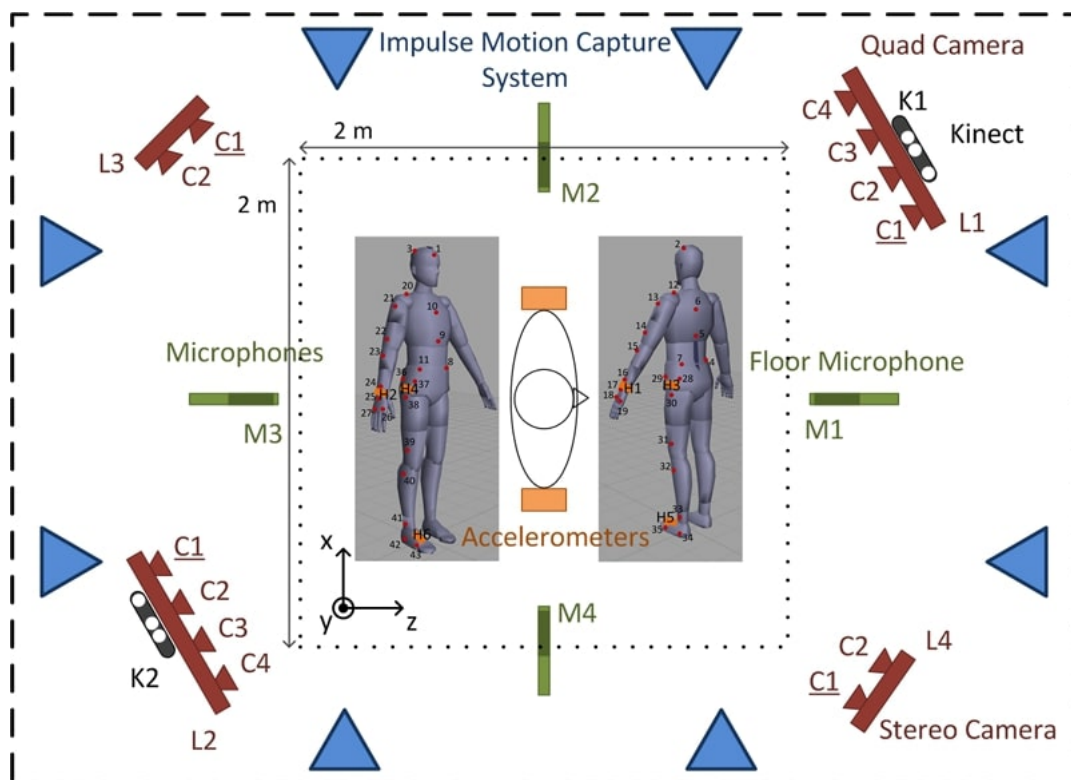


Figura 1: Sistema di cattura delle azioni [10]

Sistema di cattura del movimento È stato utilizzato il sistema Impulse (PhaseSpace Inc., San Leandro, CA), che cattura la posizione 3D di indicatori LED alla frequenza di 480 Hz. Il sistema

utilizza videocamere con risoluzione di 3600x3600 pixel, per una precisione nell'ordine del millimetro. Inoltre sono state utilizzate 8 camere per il riconoscimento del movimento, disposte in forma circolare.

Sistema di videocamere Per la cattura dei dati video multi-prospettiva sono state utilizzate 12 camere Dragonfly2 (Point Grey Research Inc., Richmond, BC, Canada), con una risoluzione di 640x480 pixel, disposte in 4 gruppi: due gruppi da 2, e due gruppi da 4. Le immagini sono state catturate nel formato RAW con un frame rate di circa 22 Hz.

Sistema Kinect Sono state posizionate due Microsoft Kinect in direzione opposta, in modo da acquisire i dati relativi alla profondità. Ogni Kinect ha acquisito immagini a colori, con risoluzione di 640x480 pixel e con frequenza di 30 Hz. Questi dati sono stati trasmessi via Bluetooth ad un computer che ha aggiunto un timestamp ad ogni frame.

Accelerometri Sono stati fatti indossare ai soggetti 6 accelerometri wireless a tre assi. Essi sono stati montati sulle caviglie, sui polsi e sui fianchi e hanno effettuato registrazioni con una frequenza di 30 Hz.

Sistema audio Sono stati utilizzati 4 microfoni, posizionati attorno alla zona di azione, come mostrato in figura 1. Tre dei microfoni erano posizionati su treppiedi, mentre il restante è stato ancorato al terreno, in modo da catturare i suoni generati a livello del pavimento. La registrazione audio è stata effettuata con una frequenza di 48 kHz attraverso un convertitore analogico-digitale e un amplificatore, collegato ad un registratore digitale.

Numero	Azione
1	Saltare sul posto
2	Effettuare salti "Jack"
3	Fare dei piegamenti
4	Dare dei pugni (Boxing)
5	Salutare a due mani
6	Salutare ad una mano (destra)
7	Applaudire
8	Lanciare una palla
9	Sedersi e poi rialzarsi
10	Sedersi
11	Alzarsi
12	Posa a T

Tabella 2: Le 11 azioni di Berkeley MHAD

3.3 PAMAP2

PAMAP2 (Physical Activity Monitoring for Aging People) [11] è stato sviluppato nel 2012 da Attila Reiss e dal suo gruppo di ricerca. Esso è scaricabile dal sito [12].

Questo dataset contiene dati raccolti da 9 soggetti, che hanno svolto 18 azioni ciascuno. Ogni soggetto indossava 3 unità di misura inerziale e un dispositivo per il monitoraggio del battito cardiaco. Questi soggetti, 8 maschi e una femmina, erano principalmente impiegati o studenti, con età tra i 24 e i 30 anni circa (in particolare, 27.22 ± 3.31 anni) e con indice di massa corporea (BMI) di $25.11 \pm 2.62 \text{ kg/m}^2$.

Le unità di misura inerziale sono 3 Colibri wireless IMU (Inertial Measurement Unit) [13], unità piccole e leggere contenenti ciascuna un sensore MEMS a 3 assi. Un sensore MEMS è formato da due accelerometri, un giroscopio ed un magnetometro, i quali hanno registrato dati ad una frequenza di 100Hz. Le unità sono state posizionate in 3 diverse parti del corpo del soggetto: una sul petto, una sul polso della mano dominante e una sulla caviglia della gamba del lato dominante (un solo soggetto era mancino, tutti gli altri erano destri).

Per quanto riguarda il sensore per il battito cardiaco, è stato utilizzato il BM-CS5SR della BM innovations GmbH, con una frequenza di campionamento di circa 9 Hz.

Ogni soggetto ha seguito un protocollo formato da 12 attività, alle quali sono state aggiunte delle azioni opzionali proposte al soggetto, di cui 6 sono state svolte da alcuni dei soggetti.

ID	Azione
1	Sdraiarsi
2	Sedersi
3	Alzarsi in piedi
4	Camminare
5	Correre
6	Andare in bicicletta
7	Nordic walking
9	Guardare la TV
10	Lavorare al computer
11	Guidare una macchina
12	Salire le scale
13	Scendere le scale
16	Pulire con l'aspirapolvere
17	Stirare
18	Piegare il bucato
19	Pulire la casa
20	Giocare a calcio
24	Saltare la corda

Tabella 3: Le 18 azioni di PAMAP2

Il dataset è formato da 54 colonne, ognuna delle quali è composta dai seguenti elementi: 1) *timestamp*; 2) *ID dell'attività*; 3) *battito cardiaco (bpm)*; 4-20) *dati del Colibri IMU sulla mano*; 21-37) *dati del Colibri IMU sul petto*; 38-54) *dati del Colibri IMU sulla caviglia*.

I dati dei Colibri IMU sono così formati: 1) *temperatura*; 2-4) *dati accelerometro 3D, con scala $A \pm 16g$* ; 5-7) *dati accelerometro 3D, con scala $A \pm 6g$* ; 8-10) *dati giroscopio 3D*; 11-13) *dati magnetometro 3D*; 14-17) *orientamento*.

3.4 Opportunity

Il dataset Opportunity [14], scaricabile dal sito ufficiale [15], è stato sviluppato dal gruppo di ricerca Opportunity e pubblicato nel 2011.

Per la raccolta dei dati, sono stati utilizzati sensori indossati dai soggetti, sensori su oggetti e sensori per l'ambiente. L'ambiente è un ambiente controllato, che simula un appartamento adibito a studio, con una sdraio, una cucina, un tavolo, una sedia, una macchina del caffè e delle porte che danno accesso all'esterno.

I sensori indossati sono 7 unità di misurazione inerziale, 12 sensori di accelerazione 3D e 4 sensori di localizzazione 3D. Sono stati posizionati sensori di accelerazione 3D e di velocità di rotazione 2D su ognuno dei 12 oggetti preparati ed utilizzati. Nell'ambiente sono invece stati disposti 12 interruttori e 8 sensori di accelerazione 3D.

Le 7 unità di misurazione inerziale forniscono dati relativi all'accelerazione 3D, alla velocità angolare 3D, al campo magnetico 3D e all'orientamento del sensore. 5 unità sono posizionate nella parte alta del corpo, mentre le restanti due sulle scarpe del soggetto.

I sensori di accelerazione 3D sono stati posizionati nella parte alta del corpo, sui fianchi e sulle gambe. Quelli di localizzazione, invece, sono stati posizionati nella parte anteriore e posteriore di entrambe le spalle.

I sensori sugli oggetti permettono di determinare quali oggetti sono stati utilizzati e l'utilizzo che ne è stato fatto; quelli posizionati nell'ambiente, invece, sono stati posti in cassetti, porte e apparecchiature della cucina. Gli interruttori sono stati inseriti a triplete all'interno del frigorifero, della lavastoviglie, e di due cassetti, in modo da capire se essi fossero aperti, semi-chiusi o chiusi. I sensori di accelerazione, infine, sono stati posizionati sugli utensili e gli elettrodomestici della cucina.

Sono stati coinvolti nella raccolta dati 4 soggetti, ognuno dei quali ha svolto 6 serie di attività: 5 di esse sono composte da attività di vita quotidiana ben precise (ADL run), mentre la rimanente è una ripetizione di una sequenza di attività (drill run).

La serie ADL è la seguente:

1. inizio: da seduto su una sdraio, alzarsi;
2. muoversi all'interno della stanza, controllare che tutti gli oggetti siano al loro posto, nei cassetti o sulle mensole;
3. uscire fuori e fare una passeggiata attorno all'edificio;
4. preparare un caffè con latte e zucchero, utilizzando la macchina del caffè;
5. bere a sorsi il caffè, mentre si cammina nell'ambiente;

6. preparare un panino, con pane, formaggio e salame, utilizzando il coltello per il pane e altri coltelli e piatti;
7. mangiare il panino;
8. risistemare gli oggetti al loro posto o all'interno della lavastoviglie, e pulire il tavolo;
9. break: sedersi sulla sdraio.

La sequenza drill, invece, consiste di 20 ripetizioni della seguente sequenza:

1. aprire e chiudere il frigorifero;
2. aprire e chiudere la lavastoviglie;
3. aprire e chiudere 3 cassette ad altezze diverse;
4. aprire e chiudere la porta 1;
5. aprire e chiudere la porta 2;
6. accendere e spegnere le luci;
7. pulire il tavolo;
8. bere mentre si è in piedi;
9. bere mentre si è seduti.

3.5 DSADS

DSADS (Daily and Sports Activities Data Set) [16], scaricabile dal sito [17], è un dataset sviluppato dal professor Billur Barshan, della Bilkent University, in Turchia, e pubblicato nel 2013.

Questo set comprende dati relativi a 19 attività di vita quotidiana e di sport, svolte da 8 soggetti per 5 minuti. Questi 5 minuti sono divisi in segmenti da 5 secondi.

I dati sono stati raccolti grazie all'utilizzo di cinque unità Xsens MTx, fatte indossare sul torso, sulle braccia e sulle gambe dei soggetti. Ognuna di queste è composta da un accelerometro 3D, un giroscopio 3D e da un magnetometro 3D.

Gli otto soggetti sono 4 femmine e 4 maschi di età compresa tra i 20 e i 30 anni, ognuno dei quali ha effettuato, nel proprio stile, ogni azione per 5 minuti.

Le attività (tabella 4) sono state svolte all'interno della Bilkent University Sports Hall, all'interno dell'edificio di Electrical and Electronics Engineering, e in un'area pianeggiante nel campus dell'università.

ID	Azione
A1	Stare seduti
A2	Stare in piedi
A3	Stare sdraiati sulla schiena
A4	Stare sdraiati sul lato
A5	Salire le scale
A6	Scendere le scale
A7	Stare in un ascensore fermo
A8	Muoversi all'interno dell'ascensore
A9	Camminare in un parcheggio
A10	Camminare su un tapis roulant ad una velocità di 4 km/h sul piano
A11	Camminare su un tapis roulant ad una velocità di 4 km/h, con un'inclinazione di 15 gradi
A12	Correre su un tapis roulant ad una velocità di 8 km/h
A13	Fare esercizio con uno stepper
A14	Fare esercizio su un cross trainer
A15	Allenarsi su una cyclette in posizione orizzontale
A16	Allenarsi su una cyclette in posizione verticale
A17	Fare canottaggio
A18	Saltare
A19	Giocare a basket

Tabella 4: Le 19 attività di DSADS

4 Dataset con segnali radio e WiFi

4.1 OPERAnet

OPERAnet [5], pubblicato nel 2021, è parte del progetto OPERA, ed è scaricabile dal sito [18]. Si tratta di un dataset per il riconoscimento multi-modale di attività umane. I dati sono stati raccolti da 6 soggetti che hanno svolto 6 azioni in due stanze diverse.

Numero	Attività
1	Sedersi su una sedia
2	Alzarsi in piedi dalla sedia
3	Sdraiarsi sul pavimento
4	Alzarsi in piedi dal pavimento
5	Camminare
6	Ruotare il corpo

Tabella 5: Le 6 azioni di OPERAnet

Le misurazioni sono state portate a termine grazie all'utilizzo di dispositivi a radio frequenza (RF) e sensori di visione o infrarossi, a formare 4 componenti.

Il primo è un sistema Wi-Fi CSI (channel State Information), implementato utilizzando delle Intel5300 WiFi Network Interface Card (NIC) con una configurazione MIMO a 3x3 antenne al ricevitore e al trasmettitore, agenti nella banda WiFi 5G. I dati sono stati raccolti da due ricevitori, posizionati in posizioni diverse rispetto al singolo trasmettitore.

Il secondo è un sistema radar WiFi passivo (PWR), implementato utilizzando la piattaforma USRP-2945 con 4 canali sincronizzati, comprendenti un canale di riferimento e 3 canali di sorveglianza. Ogni ricevitore, posizionato in un posto diverso rispetto agli altri, è stato equipaggiato con un'antenna direzionale da 6-dB.

Il terzo componente consiste di tre sistemi Ultra-Wideband (UWB) implementati utilizzando i kit Decawave MDEK1001 e EVK1000 UWB. Un sistema (attivo) è stato utilizzato per registrare le coordinate XY del soggetto all'interno dell'area di analisi. Altri due sistemi (passivi) sono stati posizionati all'interno dell'area di monitoraggio.

Infine, il quarto componente è un sistema di visione a infrarosso implementato utilizzando due Kinect v2 Microsoft, utilizzati per tenere traccia delle informazioni relative ai cambiamenti nel tempo dello scheletro corporeo del soggetto.

4.2 Mendeley

Il dataset Mendeley [19], pubblicato nel 2020 e scaricabile dal sito ufficiale [20], contiene dati di segnali WiFi raccolti in 3 ambienti al chiuso. Ognuno dei 30 partecipanti ha svolto delle azioni tra due dispositivi, ognuno con scheda di rete Intel 5300 installata. Al trasmettitore è stata utilizzata un'antenna, mentre al ricevitore sono state utilizzate 3 antenne.

Nel primo e nel secondo ambiente è stata utilizzata un'architettura LOS (line-of-sight), mentre nel terzo ambiente è stata utilizzata un'architettura NLOS (non-line-of-sight).

In ogni ambiente, 10 dei soggetti hanno svolto 5 esperimenti per 20 volte.

Numero	Esperimento
1	Cadere da posizione seduta
2	Cadere da posizione in piedi
3	Camminare
4	Alzarsi da una sedia e sedersi sulla sedia
5	Raccogliere una penna dal pavimento

Tabella 6: I 5 esperimenti di Mendeley

4.3 WiAR

WiAR [21] [22] [23], pubblicato nel 2017 e scaricabile dalla pagina GitHub [24], è un dataset contenente 16 attività svolte da 10 soggetti. 5 maschi e 5 femmine, in 3 ambienti diversi.

Ogni attività è stata svolta da ogni volontario 30 volte, e appartiene ad una di tre categorie: attività riguardanti la parte alta del corpo, attività riguardanti la parte bassa del corpo, attività riguardanti l'intero corpo [25].

All'interno di WiAR vengono fornite informazioni RSSI e CSI, nonché i dati grezzi raccolti. Per raccogliere tali informazioni, all'interno della stanza sono stati posizionati il soggetto e un operatore. Il trasmettitore manda 30 pacchetti al secondo. Il soggetto è collocato tra il trasmettitore e il ricevitore, e svolge le 30 ripetizione dell'azione, per poi passare alla successiva. In questo modo si possono registrare i dati RSSI e CSI relativi al segnale WiFi.

Il trasmettitore è equipaggiato con un'antenna, mentre il ricevitore ne presenta 3.

Numero	Attività
1	Saluto orizzontale
2	Saluto verticale
3	Saluto con due braccia
4	Lancio alto
5	Drwa_x
6	Scrivere una spunta
7	Lanciare della carta
8	Calcio frontale
9	Calcio laterale
10	Piegarsi
11	Sbattere le mani
12	Camminare
13	Fare una telefonata
14	Bere dell'acqua
15	Sedersi
16	Fare uno squat

Tabella 7: Le 16 azioni di WiAR

5 Dataset contenenti video

5.1 YouTube Action (UCF11)

Il dataset UCF11 (YouTube Action) [26], creato nel 2009 e scaricabile dal sito [27], è il dataset di partenza da cui sono stati ricavati, con degli ampliamenti, UCF50 e il successivo UCF101.

Esso contiene 11 categorie di azioni, per ognuna delle quali i video sono stati divisi in 25 gruppi, con più di 4 clip ciascuno.

I video contenuti nel dataset sono in totale 1600 e sono in formato mpeg4.

Numero	Attività
1	Tiro di basketball
2	Ciclismo
3	Immersione
4	Swing di golf
5	Andare a cavallo
6	Palleggiare con un pallone da calcio
7	Andare sull'altalena
8	Swing di tennis
9	Saltare sul tappeto elastico
10	Schiacciata di pallavolo
11	Portare a spasso il cane

Tabella 8: Le 11 azioni di UCF11

5.2 UCF50

UCF50 [28], scaricabile dal sito [27], è un dataset per il riconoscimento di attività umane, comprendente video realistici ricavati da YouTube appartenenti a 50 categorie di azioni. Esso è l'estensione del dataset YouTube Action (UCF11) ed è stato pubblicato nel 2012.

È stato ampliato successivamente nel dataset UCF101 e di conseguenza ne è una parte.

Per ogni azione, i video sono stati divisi in 25 gruppi, ognuno dei quali contiene più di 4 clip simili per contesto, punto di vista, etc.

Numero	Attività	Numero	Attività
1	Baseball	26	Suonare la chitarra
2	Basketball	27	Suonare il pianoforte
3	Sollevamento pesi su panca piana	28	Suonare la tabla
4	Ciclismo	29	Suonare il violino
5	Biliardo	30	Salto con l'asta
6	Nuoto a rana	31	Esercizio su cavallina
7	Sollevamento pesi	32	Pull ups (trazioni)
8	Immersione	33	Pugno di box
9	Suonare la batteria	34	Push ups (piegamenti)
10	Scherma	35	Arrampicata indoor
11	Swing di golf	36	Arrampicata sulla corda
12	Salto in alto	37	Canottaggio
13	Corsa con i cavalli	38	Piroetta della salsa
14	Andare a cavallo	39	Andare sullo skateboard
15	Fare hula hoop	40	Sciare
16	Lancio del giavellotto	41	Andare sulla moto d'acqua
17	Fare il giocoliere con delle palline	42	Palleggiare con un pallone da calcio
18	Jumping Jack	43	Andare sull'altalena
19	Saltare la corda	44	Tai Chi
20	Kayak	45	Swing di tennis
21	Affondi	46	Lancio del disco
22	Parata militare	47	Saltare sul tappeto elastico
23	Impastare	48	Schiacciata di pallavolo
24	Usare i nunchaku	49	Portare a spasso il cane
25	Stendere la pizza	50	Usare uno yo-yo

Tabella 9: Le 50 azioni di UCF50

5.3 UCF101

UCF101 [6], scaricabile dal sito ufficiale della University of Central Florida [27], è un dataset di video realistici raccolti da YouTube, comprendenti 101 azioni. È un'estensione del dataset UCF50.

Esso è stato creato dal centro di ricerca in computer vision della University of Central Florida nel 2012 e contiene 13320 video, diversi tra loro per svariate caratteristiche. Tali video puntano a rappresentare azioni realistiche, e non svolte da attori.

I video, rappresentanti 101 categorie, sono divisi, per ogni categoria, in 25 gruppi da 4-7 video, simili tra loro per sfondo simile, punto di vista simile, etc.

Le categorie di azione, invece, possono essere divise in 5 gruppi:

1. interazioni uomo-oggetto,
2. solo movimento del corpo umano,
3. interazione uomo-uomo,
4. suonare uno strumento musicale,
5. sport.

Numero	Attività	Numero	Attività
1	Mettere il makeup	52	Tiro con l'arco
2	Mettere il rossetto	53	Esercizio sulla trave
3	Asciugarsi i capelli	54	Baseball
4	Lavarsi i denti	55	Basketball
5	Tagliare in cucina	56	Schiacciata di basketball
6	Usare il martello	57	Sollevamento pesi su panca piana
7	Fare hula hoop	58	Ciclismo
8	Fare il giocoliere con delle palline	59	Biliardo
9	Saltare la corda	60	Bowling
10	Lavorare a maglia	61	Boxing con sacco
11	Impastare	62	Boxing con speed bag
12	Usare i nunchaku	63	Nuoto a rana
13	Stendere la pizza	64	Sollevamento pesi
14	Tagliarsi la barba	65	Salto da una scogliera (cliff diving)
15	Andare sullo skateboard	66	Cricket bowling
16	Palleggiare con un pallone da calcio	67	Tiro di cricket
17	Scrivere su una tastiera	68	Immersione
18	Scrivere su una lavagna	69	Scherma

Numero	Attività	Numero	Attività
19	Usare uno yo-yo	70	Rigore di hockey su prato
20	Lavare il pavimento	71	Ginnastica artistica
21	Bambino che gattona	72	Presa di un frisbee
22	Soffiare le candeline	73	Nuoto a stile libero
23	Squat	74	Swing di golf
24	Pushup in verticale	75	Lancio del martello
25	Camminare sulle mani	76	Salto in alto
26	Jumping Jack	77	Corsa con i cavalli
27	Affondi	78	Andare a cavallo
28	Pull ups (trazioni)	79	Danza sul ghiaccio
29	Push ups (piegamenti)	80	Lancio del giavellotto
30	Arrampicata indoor	81	Kayak
31	Arrampicata sulla corda	82	Salto in lungo
32	Andare sull'altalena	83	Esercizio su barre parallele
33	Tai Chi	84	Salto con l'asta
34	Saltare sul tappeto elastico	85	Esercizio su cavallina
35	Portare a spasso il cane	86	Pugno di box
36	Push ups al muro	87	Rafting
37	Marciare con una banda	88	Canottaggio
38	Tagliare i capelli	89	Lancio del peso
39	Massaggio alla testa	90	Sciare
40	Parata militare	91	Andare sulla moto d'acqua
41	Piroetta della salsa	92	Sky diving
42	Suonare la batteria	93	Rigore di calcio
43	Suonare il violoncello	94	Esercizio con anelli
44	Suonare il daf	95	Sumo wrestling
45	Suonare il dhol	96	Fare surf
46	Suonare il flauto	97	Tiro di tennistavolo
47	Suonare la chitarra	98	Swing di tennis
48	Suonare il pianoforte	99	Lancio del disco
49	Suonare il sitar	100	Esercizio su barre asimmetriche
50	Suonare la tabla	101	Schiacciata di pallavolo
51	Suonare il violino		

Tabella 10: Le 101 azioni di UCF101

Le azioni da 1 a 19 fanno parte del primo gruppo, quelle 20 a 35 del secondo, quelle da 36 a 40 del terzo, quelle da 41 a 50 al quarto e le restanti del quinto.

5.4 UCF Sports Action

UCF Sports Action [29] [30], scaricabile al sito [27], è stato creato nel 2008 dal centro di ricerca in computer vision della University of Central Florida.

Esso contiene un set di video, rappresentanti 10 tipologie di azioni raccolte da varie tipologie di sport che vengono spesso trasmessi sui canali televisivi BBC e ESPN. Le sequenze video sono state ottenute da siti di video stock, come BBC Motion Gallery e GettyImages.

Numero	Attività	Numero di video
1	Immersione	14 video
2	Swing di golf	18 video
3	Calcio ad un pallone	20 video
4	Sollevamento pesi	6 video
5	Andare a cavallo	12 video
6	Correre	13 video
7	Andare sullo skateboard	12 video
8	Esercizio su cavallina	20 video
9	Esercizio su barre	13 video
10	Camminare	22 video

Tabella 11: Le 10 azioni di UCF Sports Action

Il dataset contiene un totale di 150 sequenze video di risoluzione 720x480.

5.5 HMDB51

Il dataset HMDB51 [31], scaricabile dal sito ufficiale [32], è stato creato nel 2011 dai Serre Lab della Brown University. Esso è composto da 6849 clip, appartenenti a 51 categorie di azioni recuperate da varie fonti, quali film, YouTube, video di Google e database pubblici. Ogni categoria contiene minimo 101 clip.

Le categorie di azioni possono essere raggruppate in 5 gruppi.

Ogni clip, oltre all'etichetta relativa alla categoria di azione, è associata anche ad una meta-etichetta contenente delle informazioni sulle proprietà della clip. Un criterio definisce 3 livelli di qualità video, e ad ogni clip è associato uno di questi livelli.

Per rimuovere la variabilità derivante dal movimento della videocamera, sono state applicate ai video delle tecniche di stabilizzazione che agiscono sui frame che costituiscono la clip.

Gruppo di appartenenza	Numero	Attività
Azioni facciali generali	1	Sorridere
	2	Ridere
	3	Masticare
	4	Parlare
Azioni facciali con utilizzo di oggetti	5	Fumare
	6	Mangiare
	7	Bere
Movimenti generali del corpo	8	Fare la ruota
	9	Applaudire
	10	Arrampicata
	11	Salire le scale
	12	Immersione
	13	Cadere sul pavimento
	14	Rovescio di tennistavolo
	15	Verticale sulle mani
	16	Saltare
	17	Pull up
	18	Push up
	19	Correre
	20	Sedersi
	21	Sit up (addominali)
	22	Capriola
	23	Alzarsi in piedi
	24	Girarsi
	25	Camminare
	26	Salutare

Gruppo di appartenenza	Numero	Attività
Movimenti del corpo con utilizzo di oggetti	27	Pettinare i capelli
	28	Ricevere una palla
	29	Estrarre la spada
	30	Dribblare
	31	Golf
	32	Colpire qualcosa
	33	Calciare un pallone
	34	Raccogliere
	35	Versare
	36	Spingere qualcosa
	37	Andare in bici
	38	Andare a cavallo
	39	Tirare una palla (basket)
	40	Tirare con l'arco
	41	Sparare con una pistola
	42	Swing di baseball
	43	Esercizio con la spada
	44	Lanciare
Movimenti del corpo per interazioni con esseri umani	45	Scherma
	46	Abbracciare
	47	Calciare qualcuno
	48	Baciare
	49	Tirare un pugno
	50	Stringere la mano
	51	Combattere con la spada

Tabella 12: Le 51 azioni di HMDB51

6 Confronti tra dataset

Per valutare la qualità dei dataset presentati, e per avere una visione d'insieme del materiale analizzato, è utile effettuare vari tipi di confronti tra i dataset.

Questi confronti possono tenere conto di varie caratteristiche dei dataset: numero di elementi, tipi di dato, numero di citazioni, e così via.

6.1 Numero di elementi

Le dimensioni di un dataset ci permettono di confrontare l'importanza del lavoro svolto dai ricercatori nella creazione di esso.

La quantità di dati presenti, infatti, è indicatore del tipo di impegno e del tipo di risorse che sono state impiegate nello sviluppo del progetto.

Per i dataset contenenti video, il numero di video è un valido indicatore della sua dimensione. Per questo, nella tabella seguente sono riportati, per ogni dataset, il numero di video presenti in esso.

Dataset	Numero di video
UCF101	13320
UCF50	6676
UCF11	1600
UCF Sports Action	150
HMDB51	6849

Tabella 13: Numero di video nei vari dataset

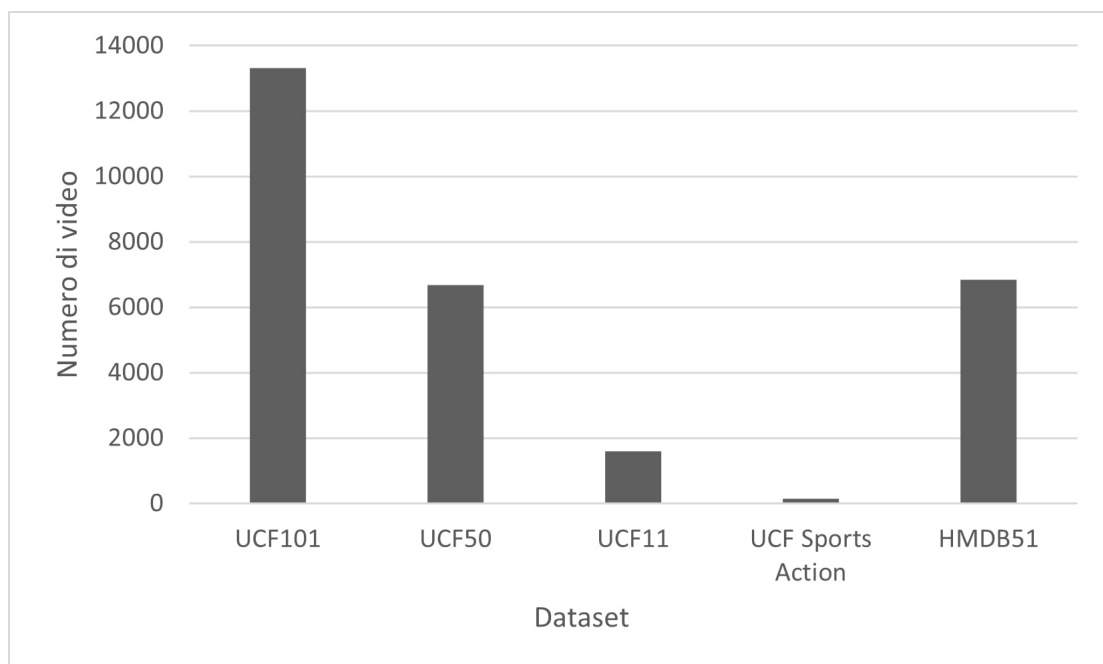


Figura 2: Grafico relativo al numero di video nei vari dataset

UCF101 è quindi il dataset video più corposo, seguito da HMDB51 e da UCF50, di dimensioni simili.

Le dimensioni dei dataset contenenti dati da sensori, invece, sono confrontabili sia come numero di righe, e quindi di record, nel dataset, sia come numero di minuti di registrazione totale (alcuni dati sono mancanti perché non presenti in rete o negli articoli che descrivono i dataset).

Dataset	Numero di righe (record)	Numero indicativo di minuti di registrazione
WISDM	15630426	2754
Berkeley MHAD	-	82
PAMAP2	3850505	600
Opportunity	2551	-
DSADS	9120	760

Tabella 14: Dimensioni dataset contenenti dati da sensori

Dai dati presenti in tabella, si può affermare che WISDM è il dataset più grande.

Berkeley MHAD ha un numero di record maggiore rispetto a DSADS ma un minutaggio totale di acquisizione inferiore. Ciò indica che la frequenza di acquisizione dei dati utilizzata in Berkeley MHAD è maggiore rispetto a quella di DSADS. Inoltre, anche il numero di sensori utilizzati in Berkeley MHAD è maggiore rispetto a quello utilizzato in DSADS.

6.2 Numero di azioni e di soggetti

Ognuno dei dataset presentati contiene dati relativi a soggetti che effettuano determinate azioni. Il numero di tali azioni dà un'idea della grandezza del dataset e, soprattutto, della diversità all'interno di esso.

Nelle prime due tipologie di dataset, ogni soggetto effettua l'azione più volte o per più tempo, mentre nei dataset video, per ogni azione sono presenti più video diversi.

Dataset	Numero di azioni registrate	Numero di soggetti
WISDM	18	51
Berkeley MHAD	11	12
PAMAP2	18	9
Opportunity	Serie ADL: 9; serie drill: 9	4
DSADS	19	8
OPERAnet	6	6
Mendeley	5	30
WiAR	16	10
UCF101	101	-
UCF50	50	-
UCF11	11	-
UCF Sports Action	10	-
HMDB51	51	-

Tabella 15: Numero di azioni nei vari dataset

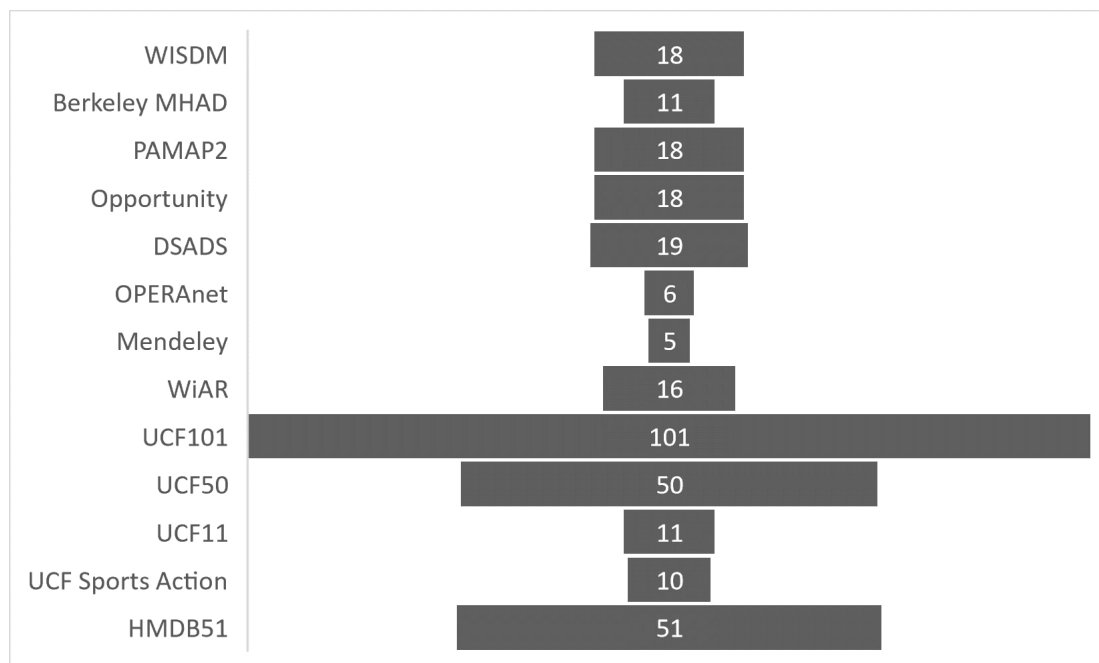


Figura 3: Grafico relativo al numero di azioni nei vari dataset

Il dataset con il maggior numero di azioni, quindi, è UCF101, seguito da HMDB51 e da UCF50. I dataset con il minor numero di azioni, invece, sono Mendeley e OPERAnet.

6.3 Numero di citazioni

Ogni dataset è legato ad un articolo scientifico. Chi crea un dataset, infatti, solitamente lo presenta in un articolo e successivamente chiede agli utilizzatori del dataset di citare tale articolo. Questo ci permette di valutare il grado di popolarità di un dataset all'interno della comunità scientifica.

Dataset	Tipologia	Articolo	Numero di citazioni
WISDM	Sensori	[4]	2848
Berkeley MHAD	Sensori	[9]	478
PAMAP2	Sensori	[11]	794
Opportunity	Sensori	[14]	665
DSADS	Sensori	[16]	596
OPERAnet	Radio e WiFi	[5]	2
Mendeley	Radio e WiFi	[19]	22
WiAR	Radio e WiFi	[21], [22], [23]	24, 6, 42
UCF101	Video	[6]	4387
UCF50	Video	[28]	680
UCF11	Video	[26]	1385
UCF Sports Action	Video	[29], [30]	1472, 227
HMDB51	Video	[31]	3222

Tabella 16: Numero di citazioni (al 27-08-2022) relative ai dataset

Nel seguente grafico, è visibile la differenza tra il numero di citazioni dei vari dataset.

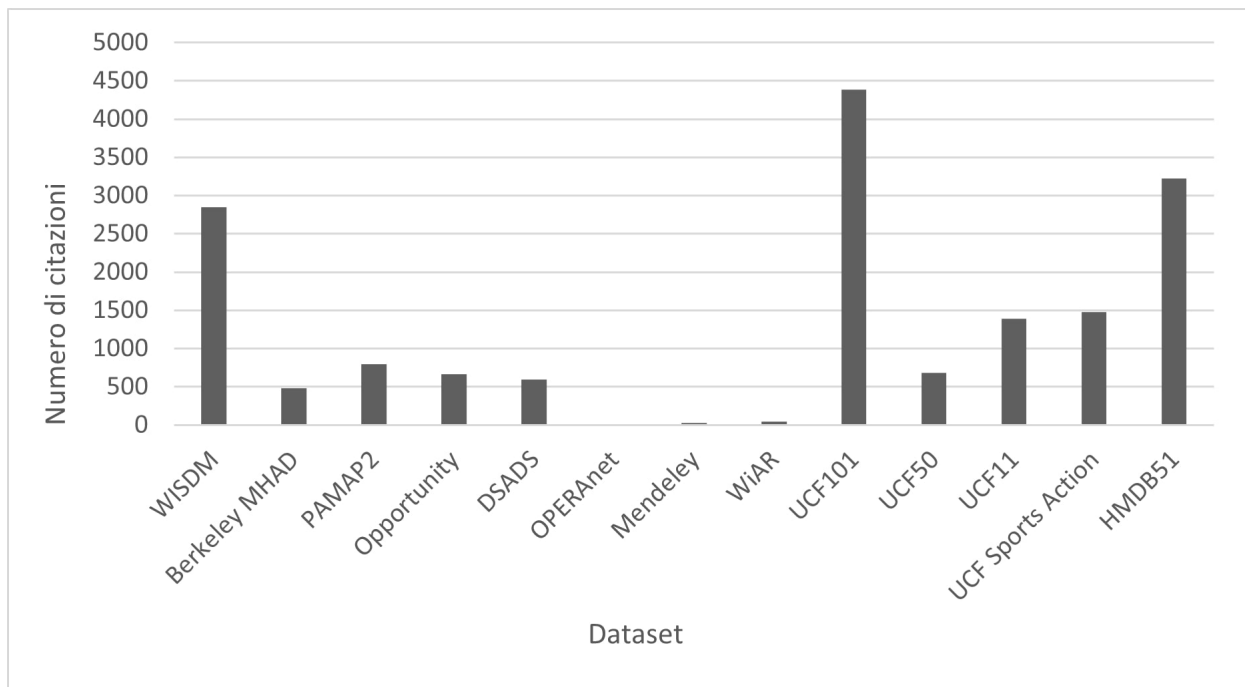


Figura 4: Grafico relativo al numero di citazioni

Un altro interessante aspetto di confronto e analisi è il numero di citazioni degli articoli che hanno utilizzato i dataset (e quindi citato l'articolo di riferimento).

Nella tabella seguente sono presentati, per ogni dataset, i 3 articoli scientifici più citati che hanno utilizzato tale dataset.

Dataset	Articolo che l'ha utilizzato	Numero di citazioni
WISDM ([4])	[33]	1732
	[34]	970
	[35]	937
Berkeley MHAD ([9])	[36]	1770
	[37]	1139
	[38]	767
PAMAP2 ([11])	[39]	1923
	[40]	1210
	[41]	872
Opportunity ([14])	[39]	1923
	[33]	1732
	[42]	1112
DSADS ([16])	[43]	686
	[44]	597
	[45]	394
OPERAnet ([5])	[46]	0
	[47]	0
Mendeley ([19])	[48]	16
	[49]	15
	[5]	2
WiAR ([23])	[50]	326
	[51]	171
	[52]	83
UCF101 ([6])	[53]	7017
	[54]	6975
	[55]	6909
UCF50 ([28])	[56]	3492
	[57]	1848
	[58]	1000
UCF11 ([26])	[54]	6975
	[59]	5699
	[6]	4387

Dataset	Articolo che l'ha utilizzato	Numero di citazioni
UCF Sports Action ([29])	[6]	4387
	[60]	3596
	[31]	3222
HMDB51 ([31])	[53]	7017
	[61]	6212
	[62]	5048

Tabella 17: Numero di citazioni (al 27-08-2022) degli articoli che hanno utilizzato i dataset

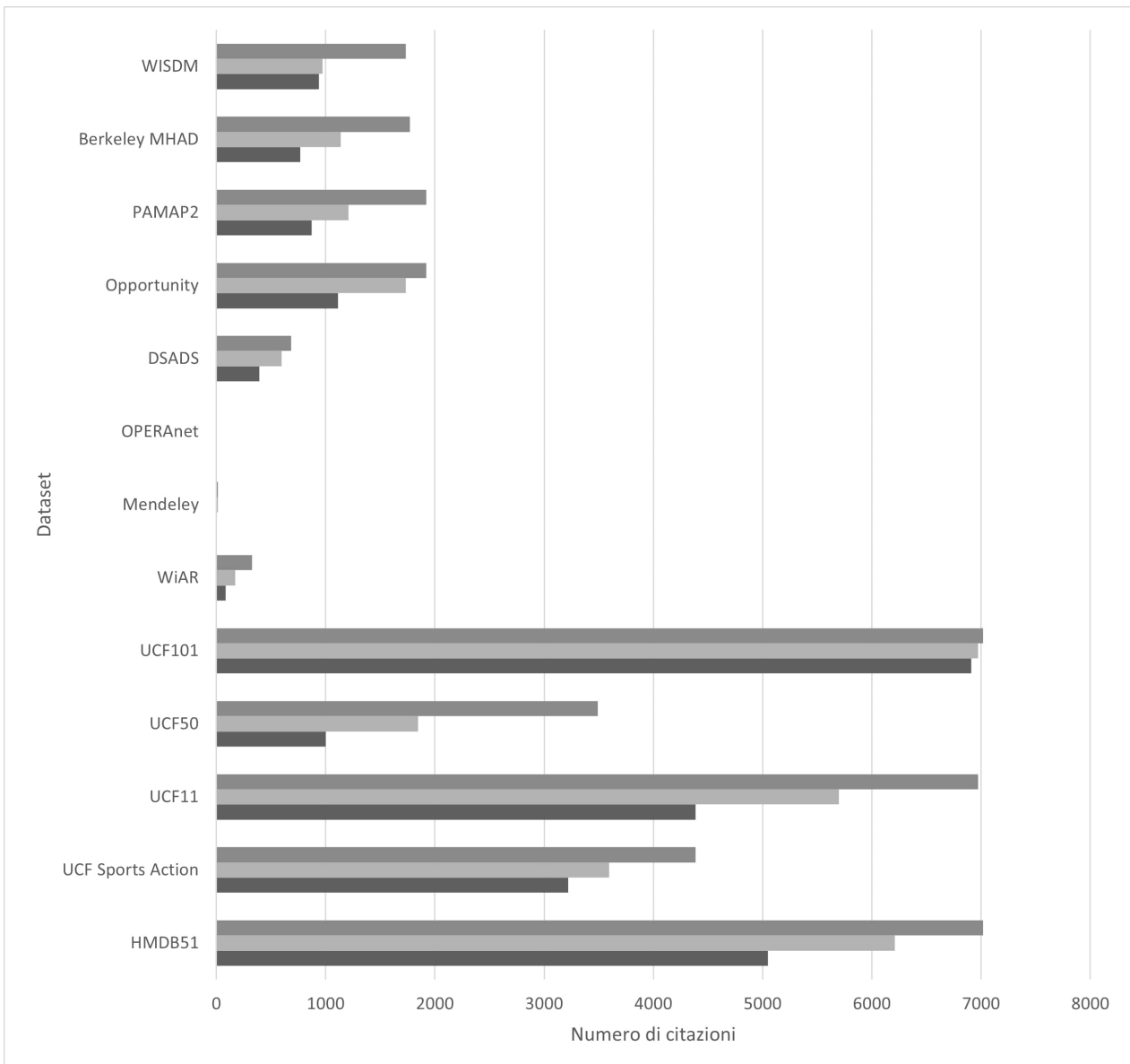


Figura 5: Grafico relativo al numero di citazioni degli articoli che hanno utilizzato i dataset

Si può quindi affermare che il dataset più popolare è UCF101, seguito da HMDB51.

In generale, si nota che i dataset video sono la tipologia di dataset più popolare, mentre, al contrario, i dataset con segnali radio e WiFi sono i meno citati e, quindi, i meno popolari.

Ogni articolo che cita uno dei dataset presentati, può utilizzare tale dataset in modo diverso.

Per analizzare in modo più approfondito gli articoli riportati in tabella 17, si possono dividere tali articoli all'interno delle seguenti categorie:

1. articoli che utilizzano il dataset originale per testare un algoritmo di riconoscimento di attività;
2. articoli che propongono una variante del dataset originale o un nuovo dataset;
3. articoli di rassegna bibliografica;
4. altro.

Di seguito, ogni articolo è associato ad una delle 4 categorie:

1. Berkeley MHAD (tutti), PAMAP2 ([39], [40]), Opportunity ([39], [42]), Mendeley ([49]), UCF101 (tutti), UCF50 ([56], [57]), UCF11 ([54]), UCF Sports Action ([60]), HMDB51 ([53], [61]);
2. WISDM ([33]), Opportunity ([33]), Mendeley ([5]), UCF11 ([6]), UCF Sports Action ([6], [31]), HMDB51 ([62]);
3. PAMAP2 ([41]), OPERAnet ([47]), Mendeley ([48]), WiAR ([50], [51]);
4. WISDM ([34], [35]), DSADS (tutti), OPERAnet ([46]), WiAR([52]), UCF50 ([58]), UCF11 ([59]).

Nel seguente grafico è rappresentato, per ogni categoria, il numero di articoli che appartengono ad essa.

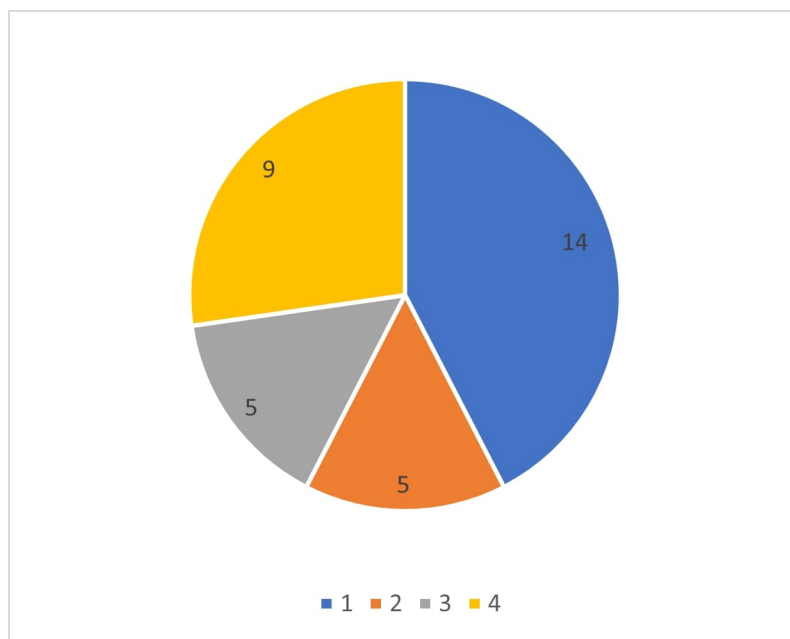


Figura 6: Grafico relativo al numero di articoli per ogni categoria

Nel seguente grafico, invece, per ogni categoria è rappresentato il numero totale delle citazioni, calcolato sommando il numero di citazioni di tutti gli articoli assegnati a quella categoria.

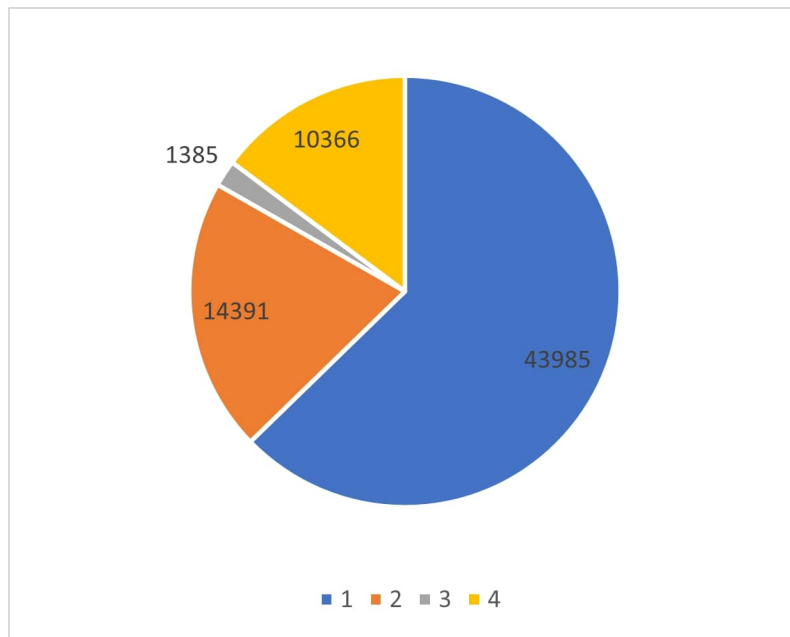


Figura 7: Grafico relativo al numero totale di citazioni per ogni categoria

È possibile notare come le proporzioni cambino se si considera il numero di articoli oppure il numero di citazioni totali.

7 Conclusioni

Nella tesina sono stati descritti e confrontati i dataset aperti più citati in letteratura nel campo del riconoscimento di attività umane.

Si è potuta notare la varietà delle tipologie di dataset e come, anche all'interno della stessa categoria, i dataset possano essere molto diversi tra loro, sia come dimensioni che come metodo di raccolta dei dati e, soprattutto, come popolarità e importanza all'interno della comunità scientifica.

Tale varietà aumenta le possibilità di scelta a disposizione della comunità scientifica e, assieme ai numerosi dataset aperti non analizzati nel lavoro, crea un ricco substrato per la ricerca nel campo della HAR.

La tipologia dei dataset con segnali radio e WiFi non è ancora popolare, ma è attualmente un ambito di ricerca che verrà sviluppato e potrebbe assumere un livello di importanza notevole.

Nel frattempo, i dataset aperti nel campo della HAR sono molti e hanno caratteristiche diverse tra di loro, quindi il materiale per lo sviluppo di algoritmi per il riconoscimento di attività è presente e facilmente accessibile.

Riferimenti bibliografici

- [1] Florenc Demrozi, Graziano Pravadelli, Azra Bihorac e Parisa Rashidi. «Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey». In: *IEEE Access* 8 (2020), pp. 210816–210836.
- [2] Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu e Yunhao Liu. «Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities». In: *ACM Computing Surveys (CSUR)* 54.4 (2021), pp. 1–40.
- [3] Djamila Romaiissa Beddiar, Brahim Nini, Mohammad Sabokrou e Abdenour Hadid. «Vision-based human activity recognition: a survey». In: *Multimedia Tools and Applications* 79.41 (2020), pp. 30509–30555.
- [4] Jennifer R Kwapisz, Gary M Weiss e Samuel A Moore. «Activity recognition using cell phone accelerometers». In: *ACM SigKDD Explorations Newsletter* 12.2 (2011), pp. 74–82.
- [5] Mohammud J Bocus, Wenda Li, Shelly Vishwakarma, Roget Kou, Chong Tang, Karl Woodbridge, Ian Craddock, Ryan McConville, Raul Santos-Rodriguez, Kevin Chetty et al. «OPERANet: A Multimodal Activity Recognition Dataset Acquired from Radio Frequency and Vision based Sensors». In: *arXiv preprint arXiv:2110.04239* (2021).
- [6] Khurram Soomro, Amir Roshan Zamir e Mubarak Shah. «UCF101: A dataset of 101 human actions classes from videos in the wild». In: *arXiv preprint arXiv:1212.0402* (2012).
- [7] *Sito web WISDM*. URL: <https://www.cis.fordham.edu/wisdm/dataset.php>.
- [8] Gary M Weiss. «Wisdm smartphone and smartwatch activity and biometrics dataset». In: *UCI Machine Learning Repository: WISDM Smartphone and Smartwatch Activity and Biometrics Dataset Data Set 7* (2019), pp. 133190–133202.
- [9] Ferda Ofli, Rizwan Chaudhry, Gregorij Kurillo, René Vidal e Ruzena Bajcsy. «Berkeley mhad: A comprehensive multimodal human action database». In: (2013), pp. 53–60.
- [10] *Sito web Berkeley MHAD*. URL: https://tele-immersion.citris-uc.org/berkeley_mhad#dl.
- [11] Attila Reiss e Didier Stricker. «Introducing a new benchmarked dataset for activity monitoring». In: *2012 16th international symposium on wearable computers*. IEEE. 2012, pp. 108–109.
- [12] *Sito web dove scaricare il dataset PAMAP2*. URL: <https://archive.ics.uci.edu/ml/datasets/pamap2+physical+activity+monitoring>.
- [13] *Trivisio web site*. URL: <https://www.trivisio.com/inertial-motion-tracking>.
- [14] Daniel Roggen, Alberto Calatroni, Mirco Rossi, Thomas Holleczeck, Kilian Förster, Gerhard Tröster, Paul Lukowicz, David Bannach, Gerald Pirkl, Alois Ferscha et al. «Collecting complex activity datasets in highly rich networked sensor environments». In: *2010 Seventh international conference on networked sensing systems (INSS)*. IEEE. 2010, pp. 233–240.

- [15] *Sito web Opportunity*. URL: <http://www.opportunity-project.eu/index.html>.
- [16] Kerem Altun, Billur Barshan e Orkun Tunçel. «Comparative study on classifying human activities with miniature inertial and magnetic sensors». In: *Pattern Recognition* 43.10 (2010), pp. 3605–3620.
- [17] *Sito web dove scaricare il dataset DSADS*. URL: <https://archive.ics.uci.edu/ml/datasets/daily+and+sports+activities>.
- [18] *Sito dove scaricare OPERAnet*. URL: <https://figshare.com/s/c774748e127dcdecc667>.
- [19] Rami Alazrai, Ali Awad, Alsaify Baha’A, Mohammad Hababeh e Mohammad I Daoud. «A dataset for Wi-Fi-based human-to-human interaction recognition». In: *Data in brief* 31 (2020), p. 105668.
- [20] *Sito web del dataset Mendeley*. URL: <https://data.mendeley.com/datasets/v38wjmz6f6>.
- [21] Linlin Guo, Lei Wang, Jialin Liu e Wei Zhou. «A survey on motion detection using WiFi signals». In: *2016 12th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*. IEEE. 2016, pp. 202–206.
- [22] Linlin Guo, Lei Wang, Jialin Liu, Wei Zhou, Bingxian Lu Tao Liu, Guangxu Li e Chen Li. «A novel benchmark on human activity recognition using WiFi signals». In: *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE. 2017, pp. 1–6.
- [23] Linlin Guo, Lei Wang, Jialin Liu, Wei Zhou e Bingxian Lu. «HuAc: Human activity recognition using crowdsourced WiFi signals and skeleton data». In: *Wireless Communications and Mobile Computing 2018* (2018).
- [24] *WiAR GitHub page*. URL: <https://github.com/linteresa/WiAR>.
- [25] Linlin Guo, Lei Wang, Chuang Lin, Jialin Liu, Bingxian Lu, Jian Fang, Zhonghao Liu, Zeyang Shan, Jingwen Yang e Silu Guo. «Wiar: A public dataset for wifi-based activity recognition». In: *IEEE Access* 7 (2019), pp. 154935–154945.
- [26] Jingen Liu, Jiebo Luo e Mubarak Shah. «Recognizing realistic actions from videos “in the wild”». In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2009, pp. 1996–2003.
- [27] *Sito web UCF*. URL: <https://www.crcv.ucf.edu/>.
- [28] Kishore K Reddy e Mubarak Shah. «Recognizing 50 human action categories of web videos». In: *Machine vision and applications* 24.5 (2013), pp. 971–981.
- [29] Mikel D Rodriguez, Javed Ahmed e Mubarak Shah. «Action mach a spatio-temporal maximum average correlation height filter for action recognition». In: *2008 IEEE conference on computer vision and pattern recognition*. IEEE. 2008, pp. 1–8.
- [30] Khurram Soomro e Amir R Zamir. «Action recognition in realistic sports videos». In: *Computer vision in sports*. Springer, 2014, pp. 181–208.

- [31] Hildegard Kuehne, Hueihan Jhuang, Estíbaliz Garrote, Tomaso Poggio e Thomas Serre. «HM DB: a large video database for human motion recognition». In: *2011 International conference on computer vision*. IEEE. 2011, pp. 2556–2563.
- [32] *Sito web di Serre Lab*. URL: <https://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database>.
- [33] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra Perez e Jorge Luis Reyes Ortiz. «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, pp. 437–442.
- [34] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra e Jorge L Reyes-Ortiz. «Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine». In: *International workshop on ambient assisted living*. Springer. 2012, pp. 216–223.
- [35] Charissa Ann Ronao e Sung-Bae Cho. «Human activity recognition with smartphone sensors using deep learning neural networks». In: *Expert systems with applications* 59 (2016), pp. 235–244.
- [36] Yong Du, Wei Wang e Liang Wang. «Hierarchical recurrent neural network for skeleton based action recognition». In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1110–1118.
- [37] Jun Liu, Amir Shahroudy, Dong Xu e Gang Wang. «Spatio-temporal lstm with trust gates for 3d human action recognition». In: *European conference on computer vision*. Springer. 2016, pp. 816–833.
- [38] Wentao Zhu, Cuiling Lan, Junliang Xing, Wenjun Zeng, Yanghao Li, Li Shen e Xiaohui Xie. «Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks». In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 30. 1. 2016.
- [39] Francisco Javier Ordóñez e Daniel Roggen. «Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition». In: *Sensors* 16.1 (2016), p. 115.
- [40] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel e Xiaowei Xu. «DBSCAN revisited, revisited: why and how you should (still) use DBSCAN». In: *ACM Transactions on Database Systems (TODS)* 42.3 (2017), pp. 1–21.
- [41] Nils Y Hammerla, Shane Halloran e Thomas Plötz. «Deep, convolutional, and recurrent models for human activity recognition using wearables». In: *arXiv preprint arXiv:1604.08880* (2016).
- [42] Jianbo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li e Shonali Krishnaswamy. «Deep convolutional neural networks on multichannel time series for human activity recognition». In: *Twenty-fourth international joint conference on artificial intelligence*. 2015.
- [43] Ferhat Attal, Samer Mohammed, Mariam Dedabrishvili, Faicel Chamroukhi, Latifa Oukhellou e Yacine Amirat. «Physical human activity recognition using wearable sensors». In: *Sensors* 15.12 (2015), pp. 31314–31338.

- [44] Mostafa Hagi, Kerstin Thurow e Regina Stoll. «Wearable devices in medical internet of things: scientific research and commercially available devices». In: *Healthcare informatics research* 23.1 (2017), pp. 4–15.
- [45] Ran Nathan, Orr Spiegel, Scott Fortmann-Roe, Roi Harel, Martin Wikelski e Wayne M Getz. «Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures». In: *Journal of Experimental Biology* 215.6 (2012), pp. 986–996.
- [46] Robert J Piechocki, Xiaoyang Wang e Mohammad J Bocus. «Multimodal sensor fusion in the latent representation space». In: *arXiv preprint arXiv:2208.02183* (2022).
- [47] Bo Tan e Bo Sun. «Using Wi-Fi Signal as Sensing Medium: Passive Radar, Channel State Information and Followups». In: *arXiv preprint arXiv:2201.03005* (2022).
- [48] Isura Nirmal, Abdelwahed Khamis, Mahbub Hassan, Wen Hu e Xiaoqing Zhu. «Deep learning for radio-based human sensing: Recent advances and future directions». In: *IEEE Communications Surveys & Tutorials* 23.2 (2021), pp. 995–1019.
- [49] Rami Alazrai, Mohammad Hababeh, Alsaify Baha’A, Mostafa Z Ali e Mohammad I Daoud. «An end-to-end deep learning framework for recognizing human-to-human interactions using Wi-Fi signals». In: *IEEE Access* 8 (2020), pp. 197695–197710.
- [50] Yongsun Ma, Gang Zhou e Shuangquan Wang. «WiFi sensing with channel state information: A survey». In: *ACM Computing Surveys (CSUR)* 52.3 (2019), pp. 1–36.
- [51] L Minh Dang, Kyungbok Min, Hanxiang Wang, Md Jalil Piran, Cheol Hee Lee e Hyeonjoon Moon. «Sensor-based and vision-based human activity recognition: A comprehensive survey». In: *Pattern Recognition* 108 (2020), p. 107561.
- [52] Wenwen Li, Michael Batty e Michael F Goodchild. *Real-time GIS for smart cities*. 2020.
- [53] Karen Simonyan e Andrew Zisserman. «Two-stream convolutional networks for action recognition in videos». In: *Advances in neural information processing systems* 27 (2014).
- [54] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar e Li Fei-Fei. «Large-scale video classification with convolutional neural networks». In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2014, pp. 1725–1732.
- [55] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani e Manohar Paluri. «Learning spatiotemporal features with 3d convolutional networks». In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 4489–4497.
- [56] Heng Wang e Cordelia Schmid. «Action recognition with improved trajectories». In: *Proceedings of the IEEE international conference on computer vision*. 2013, pp. 3551–3558.
- [57] Heng Wang, Alexander Kläser, Cordelia Schmid e Cheng-Lin Liu. «Dense trajectories and motion boundary descriptors for action recognition». In: *International journal of computer vision* 103.1 (2013), pp. 60–79.

- [58] Hirokazu Kato, Mark Billinghurst, Ivan Poupyrev, Kenji Imamoto e Keihachiro Tachibana. «Virtual object manipulation on a table-top AR environment». In: *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*. Ieee. 2000, pp. 111–119.
- [59] Shuiwang Ji, Wei Xu, Ming Yang e Kai Yu. «3D convolutional neural networks for human action recognition». In: *IEEE transactions on pattern analysis and machine intelligence* 35.1 (2012), pp. 221–231.
- [60] Limin Wang, Yu Qiao e Xiaoou Tang. «Action recognition with trajectory-pooled deep convolutional descriptors». In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 4305–4314.
- [61] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko e Trevor Darrell. «Long-term recurrent convolutional networks for visual recognition and description». In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 2625–2634.
- [62] Joao Carreira e Andrew Zisserman. «Quo vadis, action recognition? a new model and the kinetics dataset». In: *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 6299–6308.