

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA
DIPARTIMENTO DI
INGEGNERIA DELL'ENERGIA ELETTRICA E DELL'INFORMAZIONE
"GUGLIELMO MARCONI"

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA BIOMEDICA

**RICONOSCIMENTO AUTOMATICO DEL CAMMINO IN
SOGGETTI ANZIANI TRAMITE SENSORI INERZIALI IN
CONDIZIONI DI VITA REALE**

Tesi in
Bioingegneria della Riabilitazione

Relatore:
Prof. Lorenzo Chiari
Correlatore:
Luca Palmerini

Presentata da:
Stefania Giordano

Anno Accademico 2020-2021

Abstract

Gait analysis is the subject of many research projects in the medical field since an accurate and reliable evaluation of gait characteristics can provide key information about people's health and can predict the onset of several neurodegenerative diseases affecting mainly the elderly population.

The purpose of this thesis is to automatically identify gait in elderly population through methods of automatic recognition and to describe gait quality with parameters derived from data acquired through wearable sensors in free-living conditions (where the subjects are free to walk in a natural way without any supervision).

First, a gait classification (gait vs other activities) was performed with three distinct approaches: machine learning, deep learning and Gait Event Detection Method (GEDM).

Three different sensor solutions (single sensor on L5, single sensor on wrist, two-sensors L5+wrist) were tested for machine learning and deep learning algorithms, while the GEDM method was used only with a single sensor on the lower back (L5). The findings show that the best performance (highest F-measure value) was achieved with the SVM machine learning algorithm. Between the two single sensor solutions the best performance was obtained by the sensor at the lower back. The results for the two-sensor solution are comparable to those obtained on L5. Second, gait analysis was carried out for GEDM and for the best classifier (both L5 and wrist). At the end of the analysis a report file showing gait characteristics was automatically generated; it can aid the clinician in his healthcare task and so contribute to improve wellbeing and quality of life.

Key words: gait analysis, feature extraction, wearable sensors, classification, elderly population, free-living.

Abstract

L'analisi del cammino è oggetto di molti progetti di ricerca in campo medico dal momento che una valutazione accurata ed affidabile delle sue caratteristiche può fornire informazioni chiave sulla salute della persona nonché predire l'insorgere di molte malattie neurodegenerative che affliggono principalmente la popolazione anziana. Lo scopo di questo lavoro di tesi è di identificare automaticamente il cammino nelle persone anziane attraverso metodi di riconoscimento automatico e descrivere la qualità del cammino con parametri derivati da dati acquisiti tramite sensori indossabili in condizioni di free-living (dove i soggetti sono liberi di camminare in maniera naturale senza alcuna supervisione).

Il primo step ha riguardato il processo di classificazione (cammino vs non-cammino) effettuato con tre distinti approcci: machine learning, deep learning e Gait Event Detection Method (GEDM). Per gli algoritmi di machine learning e deep learning sono state testate tre diverse soluzioni in termini di sensori adoperati (singolo sensore su L5, singolo sensore sul polso, doppio sensore L5+polso). Il metodo GEDM è stato utilizzato solo con la soluzione a singolo sensore su L5. I risultati mostrano che la migliore performance (valore di F-measure più alto) è stata ottenuta con l'algoritmo di machine learning SVM. Tra le due soluzioni a singolo sensore, la migliore performance è stata ottenuta adoperando il sensore su L5. I risultati ottenuti con la soluzione a doppio sensore sono comparabili a quelli ottenuti con la soluzione a singolo sensore su L5.

L'analisi del cammino su cui si è focalizzata la seconda parte del lavoro di tesi è stata eseguita per GEDM e per il migliore classificatore (sia per L5 che per il sensore sul polso). Al termine dell'analisi è stato generato automaticamente un report che mostra le caratteristiche del cammino e che può supportare il clinico nel prendersi cura della salute dei pazienti e pertanto contribuire a migliorarne il benessere e la qualità della vita.

Parole chiave: analisi del cammino, feature extraction, sensori indossabili, classificazione, popolazione anziana, free-living.

Indice

Introduzione	1
1 IL CAMMINO	2
1.1 Il cammino come biomarker dello stato di salute in soggetti anziani	2
1.2 Analisi del cammino	3
1.2.1 Laboratorio di analisi del cammino	4
1.2.2 Sensori indossabili	5
2 MATERIALI E METODI	7
2.1 Obiettivo del progetto	7
2.2 Dataset	7
2.3 Collocazione delle unità di misura inerziali	9
3 CLASSIFICAZIONE CAMMINO VS NON-CAMMINO	11
3.1 Recenti avanzamenti nell'ambito della classificazione di attività fisica .	11
3.2 Riconoscimento del cammino	13
3.3 Machine Learning applicato al riconoscimento del cammino	15
3.3.1 Feature Selection	17
3.3.1.1 Correlation Feature Selection (CFS)	17
3.3.1.2 Fast Correlated Based Filter (FCBF)	18
3.3.1.3 ReliefF	18
3.3.2 Sliding Window	18
3.3.3 Algoritmi di Machine Learning testati	19
3.3.3.1 K-Nearest Neighbour (K-NN)	19
3.3.3.2 Decision Tree	20
3.3.3.3 Support Vector Machine (SVM)	21
3.3.3.4 Random Forest	22
3.3.3.5 Logistic Regression	23
3.4 Deep Learning applicato al riconoscimento del cammino	24
3.4.1 LSTM	24
3.5 Valutazione delle performance degli algoritmi di classificazione	26
3.6 Gait Event Detection Method-GEDM	31

3.7 Gaitpy	35
3.8 Confronto delle performance di classificazione	36
4 QUANTIFICAZIONE E VALUTAZIONE DEL CAMMINO	38
4.1 Dal riconoscimento alla caratterizzazione del cammino	38
4.2 Identificazione degli intervalli (bout) di cammino	38
4.3 Parametri che descrivono il cammino	40
4.3.1 Stima dei parametri del cammino con il sensore sulla schiena (L5)	42
4.3.2 Stima dei parametri del cammino con il sensore sul polso	48
4.4 Report	52
4.4.1 Approfondimento sui parametri del cammino	55
5 CONCLUSIONI	58
5.1 Risultati	58
5.2 Discussione	59
5.3 Prospettive future	60
Appendice	62
Bibliografia	67

Introduzione

Il mondo sta rapidamente invecchiando, l'OMS stima che il numero di over 60 raddoppierà da qui a 30 anni e nel 2050 questa fascia d'età rappresenterà il 22% della popolazione mondiale. Per questa data la popolazione anziana supererà quella dei bambini (0-14) per la prima volta nella storia dell'umanità [1][2].

L'invecchiamento della popolazione rappresenta contemporaneamente un trionfo e una sfida per la società [3]. L'aspettativa di vita più lunga, importante traguardo raggiunto grazie a migliori condizioni igieniche e sanitarie e ai progressi tecnologici in ambito medico, non si associa necessariamente ad una migliore qualità della vita. Di fatto, aumenta sempre più il peso delle patologie cronico-degenerative legate all'invecchiamento, con conseguente crescita dei costi assistenziali e difficoltà di sostenibilità dell'intero sistema socio-sanitario. La popolazione anziana, dunque, a causa di cambiamenti fisiologici che accompagnano l'invecchiamento, è la più colpita ed esposta alle malattie. Promuovere un invecchiamento in salute, creare ambienti favorevoli e adeguare i sistemi sanitari alle esigenze di una popolazione sempre più anziana, sono gli obiettivi che l'OMS cerca di raggiungere in tempi brevi. La cura e la promozione della salute geriatrica sono, dunque, due aspetti fondamentali a cui i professionisti della sanità devono porre attenzione. L'Assemblea Generale delle Nazioni Unite ha dichiarato il periodo che ci apprestiamo a vivere, 2021-2030, *“il decennio per la vecchiaia in buona salute”* [4].

L'attività fisica è un elemento chiave per il raggiungimento di tale obiettivo per la sua capacità di preservare l'indipendenza funzionale in età avanzata e garantire una buona qualità della vita. Promuovere l'attività fisica, intesa come movimento prodotto dai muscoli scheletrici che richiede un dispendio di energia, e prevenire le cadute sono infatti tra i cinque interventi di priorità da attuare in ambito sanitario secondo quanto emerge da uno studio condotto dall'OMS [4]. Al fine di migliorare la salute cardiorespiratoria e muscolare, ridurre il rischio di malattie croniche, depressione e declino cognitivo, gli adulti over 65 anni dovrebbero svolgere almeno 150 minuti alla settimana di attività fisica aerobica di moderata intensità [5]. Oltre a promuovere l'attività fisica è importante monitorare e valutarne l'efficacia non solo per escludere controindicazioni relative o assolute ma soprattutto per identificare i soggetti a rischio da indirizzare verso una valutazione specialistica.

Capitolo 1: Il cammino

1.1 Il cammino come biomarker dello stato di salute in soggetti anziani

Tra le varie attività fisiche il cammino è considerato la forma di esercizio più facilmente praticabile poiché non richiede *training*, abilità specifiche o elevati livelli di forma fisica. È un'attività che può essere svolta a costo zero, senza alcun vincolo di orari o luoghi e che non necessita di particolari attrezzature.

Gli esperti sostengono che il cammino sia l'attività più adatta agli anziani, rappresentando un efficace mezzo per ridurre il rischio di condizioni croniche e migliorare la salute generale. Tra i benefici ad esso legati vi è il miglioramento della salute cardiovascolare con riduzione dei livelli di colesterolo, pressione sistolica e rischio di cardiopatia coronarica nonché riduzione del dolore legato a condizioni croniche-infiammatorie come l'artrite [6].

Lo studio della locomozione può rappresentare un importante mezzo per la valutazione del rischio di cadute negli anziani. Sebbene la caduta sia un evento improvviso, la modificazione di alcuni fattori che si verifica giorni o mesi prima che l'evento accada può essere associata al rischio di caduta. La velocità del cammino (*walking speed*), già dimostratasi parametro in grado di predire l'aspettativa di vita negli anziani [7], si è rivelata fondamentale anche nel predire il rischio di caduta. È stato infatti dimostrato che il suo valor medio subisce un significativo declino nei tre mesi che precedono l'evento di caduta [8].

L'analisi del cammino rappresenta, inoltre, un potente strumento che permette di identificare gli esordi di alcune malattie, monitorarne l'evoluzione, attuare terapie riabilitative personalizzate e controllarne l'efficacia per eventualmente adeguare il piano terapeutico all'evoluzione delle malattie stesse. Ciò vale in modo particolare per malattie neurodegenerative quali Parkinson e Alzheimer per le quali i disturbi del cammino sono sintomi molto invalidanti che determinano una riduzione dell'autonomia e un peggioramento della qualità della vita. Le patologie neurodegenerative sono già piuttosto diffuse ma la loro incidenza è destinata a crescere sempre più, in modo particolare nella popolazione anziana. La possibilità di effettuare una diagnosi precoce è quindi

di fondamentale importanza in quanto consente di rallentare la progressione della malattia, ridurre il dosaggio farmacologico e impostare uno schema terapeutico il più efficace e individualizzato possibile.

Uno studio recente sull'Alzheimer ha evidenziato che il modo di camminare può modificarsi diversi anni prima di sviluppare la demenza [9][10]. Questo perché la demenza è associata alla morte delle cellule cerebrali che può compromettere molti aspetti della vita quotidiana, tra cui il cammino. Gli attuali metodi di diagnosi si basano sull'osservazione e la segnalazione di sintomi cruciali e per rafforzare la fiducia nella diagnosi sono raccomandate scansioni cerebrali [9][10]. Tuttavia, questo approccio si basa su sintomi già evidenti, mentre i metodi oggettivi per sostenere la diagnosi precoce, come ad esempio valutazioni sul cammino, possono rivelare problemi di fondo prima che tali sintomi siano visibili.

Il cammino è dunque un importante strumento di misura dello stato di salute, predittore del declino cognitivo e del rischio di caduta, nonché della qualità della vita e della longevità [11]. L'analisi dei pattern del cammino può fornire informazioni significative correlate alle funzioni fisiche e neurologiche e può contribuire alla diagnosi di disordini motori in condizioni patologiche. Il suo studio è quindi un aspetto fondamentale da approfondire.

1.2 Analisi del cammino

L'analisi del cammino, parte della più generale analisi strumentale del movimento, è lo studio sistematico della locomozione e della deambulazione. Tale analisi è finalizzata a fornire risposte a precise domande cliniche correlate a specifiche decisioni cliniche e/o al monitoraggio dell'evoluzione del paziente. La valutazione dell'analisi del cammino rappresenta un valido strumento diagnostico e prognostico che, assieme all'esperta valutazione clinica, permette di fare considerazioni pretrattamento, influenzare la pianificazione della chirurgia funzionale, attuare strategie riabilitative opportune e in generale supportare decisioni in linea con gli obiettivi clinici individuati [12].

1.2.1 Laboratorio di analisi del cammino

L'analisi del cammino viene ad oggi svolta, nella stragrande maggioranza dei casi, nei laboratori di analisi del movimento, strutture dotate di complesse attrezzature quali: pedane di forza (per la misura delle reazioni fra piede e suolo), elettromiografi (per le misure dell'attività muscolare) e sistemi ottici di *motion-capture* (per le misure cinematiche) che rappresentano il gold standard [13]. Adoperare tale strumentazione rende possibile effettuare una valutazione dei parametri spaziotemporali del passo, della cinematica e della dinamica del cammino, oltre che ottenere informazioni riguardanti il dispendio energetico che il cammino comporta nel caso in cui il laboratorio sia dotato di un sistema di misura del consumo di ossigeno. Nonostante un ambiente così strutturato e strumentalizzato sia in grado di fornire misure quantitative ed oggettive vi sono anche notevoli svantaggi nell'adoperare un sistema di *gait analysis* implementato in laboratorio. Il primo aspetto negativo è legato all'elevato numero di profili professionali che esso richiede: fisiatristi, ortopedici, neurologi, medici dello sport, terapisti della neuro e psicomotricità dell'età evolutiva, nonché bioingegneri con competenze specifiche sull'apparato locomotore umano e sull'utilizzo delle strumentazioni base di un laboratorio di analisi del cammino. I professionisti che operano in un laboratorio di analisi del movimento devono possedere conoscenze in materia di biomeccanica e neurofisiologia del movimento umano in condizioni fisiologiche e patologiche, conoscenza delle potenzialità e dei limiti delle diverse tecniche di analisi utilizzabili, adeguate competenze nell'esecuzione dell'esame e nella sintesi e rielaborazione dei dati [12]. Le attrezzature presenti nel laboratorio sono costose e i tempi necessari ad equipaggiare adeguatamente i soggetti e istruirli circa il protocollo da eseguire sono generalmente lunghi. Inoltre, i laboratori sono in grado di catturare solo un breve *snapshot* della performance motoria, essendo l'analisi circoscritta ad un numero limitato di passi e dunque si configurano come ambienti non adatti ad un monitoraggio di più lungo periodo. Altra problematica importante è legata alla limitata validità ecologica del sistema, ovvero l'impossibilità di replicare l'ambiente naturale in cui l'attività fisica di una persona normalmente avviene; ciò può comportare, quindi, una modificazione dei parametri caratteristici della locomozione. Ultimo, ma non per importanza, quello descritto è un ambiente controllato dove l'attività avviene sotto stretta supervisione del personale addetto; condizione che

determina l'insorgenza di fattori psicologici che contribuiscono a creare la differenza tra il cammino in laboratorio e quello della vita quotidiana.

1.2.2 Sensori indossabili

I sensori indossabili sono dispositivi elettronici che possono, come è facilmente intuibile dal nome, essere indossati su diverse parti del corpo al fine di registrare in tempo reale i segnali che esso produce. I recenti progressi in ambito elettronico hanno incoraggiato ricercatori e scienziati ad incorporare tale tecnologia nei sistemi di salute personale. Nonostante la convenzionale analisi del cammino richieda un laboratorio di analisi del movimento dotato di un sistema stereofotogrammetrico, numerosi limiti ne riducono il potenziale utilizzo: spazio ridotto dell'ambiente di laboratorio, impossibilità di valutare numerosi cicli del passo, incompatibilità di utilizzo in condizioni di vita reale.

Per tale ragione, sta prendendo piede l'utilizzo dei sensori indossabili come strumento utile all'analisi del cammino [14]. Il successo di tali dispositivi è legato ai numerosi punti di forza che li caratterizzano:

- 1) Ridotte dimensioni;
- 2) Costi contenuti;
- 3) Monitoraggio di lungo periodo;
- 4) Collezione di dati in condizioni di vita reale;
- 5) Possibilità di utilizzo durante le Attività della Vita Quotidiana (ADL).

I device maggiormente utilizzati sono i sensori inerziali, dispositivi che applicano il principio di inerzia per misurare accelerazioni lineari (tramite accelerometri) o velocità angolari (tramite giroscopi). Tali sensori possono anche essere integrati in quella che viene chiamata Unità di Misura Inerziale (IMU) contenente al proprio interno un accelerometro triassiale, un giroscopio triassiale e, in alcuni casi, un magnetometro triassiale. Questi tre sensori misurano grandezze fisiche diverse che, opportunamente integrate, consentono la generazione di grandezze fisiche derivate, ad esempio posizione e orientamento del sensore, che rendono più completa la descrizione del fenomeno osservato [15].

Tale tecnologia, grazie ai progressi nel campo dei sistemi micro-elettromeccanici, ben si adatta ad essere utilizzata per un monitoraggio

continuo e di lungo periodo in condizioni non supervisionate quali sono quelle della vita quotidiana.

Tronco, pelvi, piedi, braccia e gambe sono segmenti corporei su cui spesso ricade la scelta del posizionamento delle unità di misura inerziali. La collocazione e il numero dei sensori è fortemente dipendente dal task e dall'*outcome* desiderato, per tale ragione, solo una loro adeguata combinazione permette di effettuare registrazioni di dati utili ad eseguire l'analisi del cammino.

Capitolo 2: Materiali e metodi

2.1 Obiettivo del progetto

L'obiettivo del lavoro di tesi è creare un *toolbox* che sia in grado di identificare automaticamente il cammino a partire dai dati registrati tramite IMU e fornire in uscita un *report* che possa fungere da supporto alla decisione clinica.

Il progetto ha come target la popolazione anziana per la quale è indicato un monitoraggio di lungo periodo dell'attività fisica, attuato per mezzo di sensori indossabili che garantiscono soluzioni robuste e poco costose, in grado di registrare in modo continuo i dati in condizioni di vita reale. L'analisi del cammino, a partire dai dati registrati, permette di estrapolare variabili significative che possono informare il clinico sullo stato di salute del soggetto. La consultazione del report permette di valutare la presenza di cambiamenti nel valore dei parametri del cammino ed interpretare tali variazioni come eventuali segni dell'insorgenza della malattia, nonché monitorare la progressione della malattia e adeguare il piano terapeutico alla sua evoluzione.

2.2 Dataset

I dati utilizzati per il lavoro rappresentano un subset di un dataset più ampio collezionato dal *Department of Neuroscience, Faculty of Medicine, at the Norwegian University of Science and Technology (NYTNU) Norway*, dal gruppo di ricerca *Geriatrics, Movement and Stroke* come parte del progetto ADAPT (*A personalized Fall Risk Assessment System for Promoting independent living*), a cui hanno preso parte 20 soggetti anziani di età 76.5 ± 5.6 anni [16].

Il protocollo di acquisizione dati è stato suddiviso in due fasi: ad un primo protocollo semi-strutturato supervisionato svolto in laboratorio ha fatto seguito un protocollo non supervisionato eseguito al di fuori dell'ambiente di laboratorio in condizioni di *free-living* [17]. In entrambe le sessioni ciascun partecipante, dotato di sensori inerziali posizionati su diverse parti del corpo, ha svolto una moltitudine di attività della vita quotidiana (ADL). Nel protocollo *in-lab* i soggetti, ricevute istruzioni sulle attività da eseguire, hanno svolto le

ADL in un laboratorio adibito a piccolo ambiente domestico. Nel protocollo *out-of-lab* i partecipanti, lasciato l'ambiente di laboratorio, sono tornati a casa per eseguire le ADL in modo naturale pur potendo includere nella routine attività predefinite (sedersi, alzarsi, camminare, sdraiarsi, salire e scendere le scale, ...), attività da svolgere senza alcuna istruzione e supervisione sul come e quando eseguirle. Per il protocollo non supervisionato, la durata media della registrazione per ciascun soggetto è stata di 1,5 ore e come gold standard è stata utilizzata una videocamera GoPro con frame rate di 29 fps fissata al petto e con obiettivo rivolto in basso verso i piedi.

Wearable Location

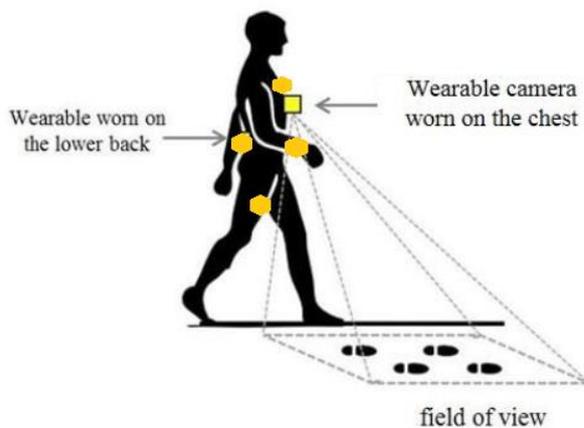


Figura 2.1: Posizionamento dei sensori e della GoPro durante il protocollo *out-of-lab*. Figura adattata da [18].

Le registrazioni video sono state poi visionate da alcuni valutatori il cui compito è stato di assegnare una *label* a ciascuna attività svolta.

Per questo lavoro di tesi è stato preso in esame il solo protocollo eseguito al di fuori dell'ambiente di laboratorio in condizioni di *free-living*.

L'analisi svolta all'interno di un ambiente controllato può, infatti, non riflettere la performance funzionale del cammino che si ha nella vita quotidiana. L'attività fisica valutata in laboratorio, in particolar modo il cammino, non rappresenta pienamente l'abilità funzionale di un soggetto in quanto egli tende a modificare il proprio comportamento quando assume la consapevolezza di essere osservato (effetto Hawthorne) [19]. Inoltre, il cammino *in-lab* riflette il cosiddetto "*can do*", la capacità, cioè quello che una persona riesce a fare in

una situazione specifica; l'“*actually do*” rappresenta, invece, la performance, ovvero ciò che la persona fa quotidianamente. La valutazione del “*can do*” potrebbe, quindi, portare a sottostimare o sovrastimare eventuali compromissioni del cammino, ragione per cui è bene rivolgere l'attenzione al cammino in condizioni di vita reale.

La scelta di utilizzare il dataset ADAPT è riconducibile a due sue peculiarità:

- 1) È uno dei più numerosi nel suo genere, ovvero acquisito in condizioni di *free-living* su soggetti anziani;
- 2) È etichettato, dunque contiene, per tutte le registrazioni, per ogni attività rilevata, le etichette di classe apposte dai valutatori; permette di valutare le performance della classificazione che verrà approfondita nei successivi capitoli.

2.3 Collocazione delle unità di misura inerziali

Un preciso e robusto posizionamento dei sensori è un requisito fondamentale per ottenere misure affidabili e ripetibili. Aspetto non meno importante nelle applicazioni da adoperare in condizioni di vita reale è l'indossabilità; ideale sarebbe poter indossare una singola IMU, che garantisca il giusto comfort a chi la indossa e contemporaneamente si dimostri una soluzione robusta.

Collocazione appropriata a tale scopo è rappresentata dalla zona lombare; il posizionamento di un sensore sulla quinta vertebra lombare (L5) è una scelta adoperata da molti ricercatori in quanto l'unità di misura, in tale posizione, risulta essere molto vicina al centro di massa e permette, dunque, di seguire in maniera più precisa il movimento del corpo. Altro punto di forza di L5 è la possibilità di fissare il sensore all'osso con semplici fasce o cinture in maniera aderente, riducendo così gli artefatti da tessuto molle sul segnale registrato.

I recenti avanzamenti tecnologici e la crescente richiesta e interesse nel monitoraggio di lungo periodo dell'attività fisica hanno spostato l'attenzione dei ricercatori su altri posizionamenti. Tra le possibili alternative si sta esplorando la collocazione di una singola IMU sul polso, segmento corporeo sul quale il sensore può essere facilmente indossato, tenuto saldamente in

posizione e ricevere, al pari di quello posizionato in zona lombare, un feedback positivo da parte dell'utente garantendo un comfort maggiore.

Inoltre, l'utilizzo di modelli astratti in cui caratteristiche nel dominio del tempo e della frequenza estratte dai dati grezzi (*raw*) registrati tramite accelerometro e giroscopio del sensore alimentano diversi modelli di machine learning (che verranno successivamente esaminati), è indipendente dall'orientazione del sensore. La non necessaria calibrazione/allineamento della IMU rende, dunque, il sensore sul polso adatto a valutazioni prolungate (*long-term*) del cammino.

Sulla base della precedente discussione, al fine di ottenere una soluzione che veda soddisfatti i requisiti di indossabilità, monitoraggio di lungo periodo e robustezza della soluzione, dei nove sensori facenti parti del progetto ADAPT, sono stati presi in esame solo i due sensori su descritti:

POSIZIONAMENTO	TIPOLOGIA	FREQUENZA CAMPIONAMENTO	SEGNALI REGISTRATI
<u>L5</u>	uSense	100 Hz	Accelerometro triassiale, giroscopio triassiale
<u>POLSO</u>	Shimmer	200 HZ	Accelerometro triassiale, giroscopio triassiale

Tabella 2.1: Descrizione dei sensori usati per l'analisi dei dati.

Capitolo 3: Classificazione cammino vs non-cammino

3.1 Recenti avanzamenti nell'ambito della classificazione di attività fisica

Gli ultimi trend mostrano un crescente interesse dei ricercatori per i sensori indossabili che consentono di acquisire dati in modo semplice ed efficiente e per le tecniche di machine learning che permettono di estrarre caratteristiche rilevanti per l'analisi [20].

Il *Machine Learning* (ML) è una branca dell'intelligenza artificiale in grado di analizzare dati e automatizzare la costruzione di modelli analitici tramite algoritmi che imparano dai dati e applicano ciò che hanno appreso per prendere decisioni informate. Gli ingressi di tali algoritmi sono le cosiddette *feature*, indici o informazioni quantitative ricavate dai dati, utili per la risoluzione del problema. I problemi di classificazione appartengono alla categoria del *machine learning supervisionato*, un approccio in cui, noto un set di esempi denominato *training-set* che include variabili di input (X) e output (Y), si cerca di fare previsioni circa il valore assunto dalla variabile di uscita di un nuovo set di dati che prende il nome di *test-set*. Nella classificazione sul dataset di esempi viene, dunque, individuato un modello capace di mettere in relazione le caratteristiche dei dati con una classe. Il modello è successivamente validato sul test-set con l'obiettivo di assegnare in maniera efficiente un'etichetta di classe ai nuovi *record*. Gli approcci di machine learning vengono adoperati per il riconoscimento dell'attività fisica in quanto sono in grado di lavorare con feature non lineari e multidimensionali.

Ghazali et al. [21] hanno identificato il cammino tra diverse attività sportive utilizzando un sensore inerziale dotato di accelerometro e giroscopio posizionato sul petto e valutando diversi algoritmi di machine learning, tra cui SVM e KNN. Zhen et al. hanno classificato 10 differenti attività fisiche (inchinarsi, camminare, salire/scendere le scale, ...) e 4 tipologie di cadute tramite MEMS e smartphone utilizzando l'algoritmo SVM [22]. Altri ricercatori hanno proposto architetture deep per il riconoscimento di attività. Una rete CNN con kernel di convoluzione modificato per adattarsi ai dati

accelerometrici registrati tramite smartphone è stata usata da *Chen e Xue* [23] per classificare 8 differenti attività svolte in un ambiente naturalistico da soggetti sani. Anche *Zebin et al.* hanno adoperato una rete CNN su dati di accelerazione e velocità angolare acquisiti tramite un set di sensori inerziali [24]. La rete CNN adoperata da *Zhao et al.* per il riconoscimento del cammino a partire da dati acquisiti con sensori indossabili ha lavorato su un segnale di input noto come Angle Embedded-Gait Dynamic Image (AE-GDI), rappresentazione dinamica bidimensionale della dinamica del cammino [25]. Reti neurali ricorrenti sono state usate per la costruzione di modelli di riconoscimento del cammino a partire da dati registrati tramite sensori indossabili [26]. Dati accelerometrici acquisiti tramite uno smartphone riposto nella tasca anteriore dei pantaloni sono stati elaborati con una rete LSTM per il riconoscimento di attività fisica [27]. *Roggen et al.* hanno eseguito il riconoscimento di molteplici attività sfruttando la rete DeepConvLSTM [28] adoperando i dati acquisiti con 7 IMU e 12 accelerometri collocati su diverse parti del corpo. Un approccio basato sull' AutoEncoder [29], rete neurale *feed-forward* costituita da uno stato di input e uno di output aventi lo stesso numero di neuroni, è stato adoperato su un dataset costituito da dati acquisiti tramite uno smartphone posto in vita per il riconoscimento di 6 differenti attività (sedersi, stare in piedi, camminare, salire/scendere le scale, sdraiarsi).

Autori	Attività riconosciute	N° e Tipologia Sensori	Posizionamento	Algoritmo
Ghazali et al.	Camminare, stare fermi, fare jogging o sprinting, saltare	1 IMU (accelerometro 3D, giroscopio 3D)	Petto	Decision tree, Discriminant Analysis, Support Vector Machine, Nearest Neighbor, Ensemble Classifier
Zhen et al.	Girarsi, alzarsi da seduto/sdraiato, inchinarsi, camminare, fare jogging, scendere e salire le scale, cadere	1 IMU (accelerometro 3D, giroscopio 3D, magnetometro 3D)	Vita	Support Vector Machine
Chen e Xue	Correre, saltare, camminare, camminare velocemente, salire e scendere le scale	Smartphone (accelerometro 3D)	Tasca dei pantaloni o di altri indumenti o Vita	Convolutional Neural Network

Zebin et al.	Camminare, salire e scendere le scale, sedersi, stare in piedi, sdraiarsi	5 IMU (accelerometro 3D, giroscopio 3D)	Pelvi, coscia dx e sx, tibia dx e sx	Convolution Neural Network
Zhao et al.	Camminare	Smartphone (accelerometro 3D, giroscopio 3D) o Smartphone+ 3IMU	Tasca dx / sx o Vita	AE-GDI + Convolutional Neural Network
Roggen et al.	Aprire e chiudere la porta, bere il caffè, mangiare, etc	7 IMU (accelerometro 3D, giroscopio 3D), 12 accelerometri 3D	Braccia, polsi, piedi, fianchi, schiena, mano, ginocchia	DeepConvLSTM
Almaskukh et al.	Sedersi, stare in piedi, camminare, scendere e salire le scale, sdraiarsi	Smartphone (accelerometro 3D, giroscopio 3D)	Vita	Stacked Auto Encoder
Murad et al.	Camminare, salire e scendere le scale, sedersi, alzarsi, sdraiarsi	Smartphone (accelerometro 3D, giroscopio 3, magnetometro 3D)	Vita	LSTM-DRNN

Tabella 3.1: Panoramica degli più recenti studi di classificazione di attività fisica.

Il progetto qui sviluppato si inserisce in questo filone di ricerca esplorando e valutando algoritmi di intelligenza artificiale e confrontandoli con metodiche alternative. Per l'esecuzione degli algoritmi implementati è stato utilizzato Matlab (Release 2021a, The MathWorks, Inc., Natick, MA, USA).

3.2 Riconoscimento del cammino

Il riconoscimento del cammino rappresenta un'area di ricerca molto attiva in ambito clinico che sempre più attira l'attenzione dei ricercatori per il suo potenziale di identificare molteplici fattori di rischio per la salute.

Il cammino non è governato unicamente dalla biomeccanica dell'apparato locomotore ma è frutto di un complesso meccanismo di coordinazione tra sistemi diversi, tra cui encefalo, midollo spinale e sistema sensoriale. L'abilità di stare in piedi e di camminare è dipendente da input visivi, vestibolari, cerebellari, motori e propriocettivi e qualsiasi problematica in uno di questi sistemi necessita di essere affrontata perché agire unicamente sulla meccanica potrebbe non essere sufficiente per la sua risoluzione. Sfida clinica attuale è riuscire a discriminare un pattern di cammino sano da uno patologico e quindi valutare la progressione dei disordini del cammino durante la locomozione.

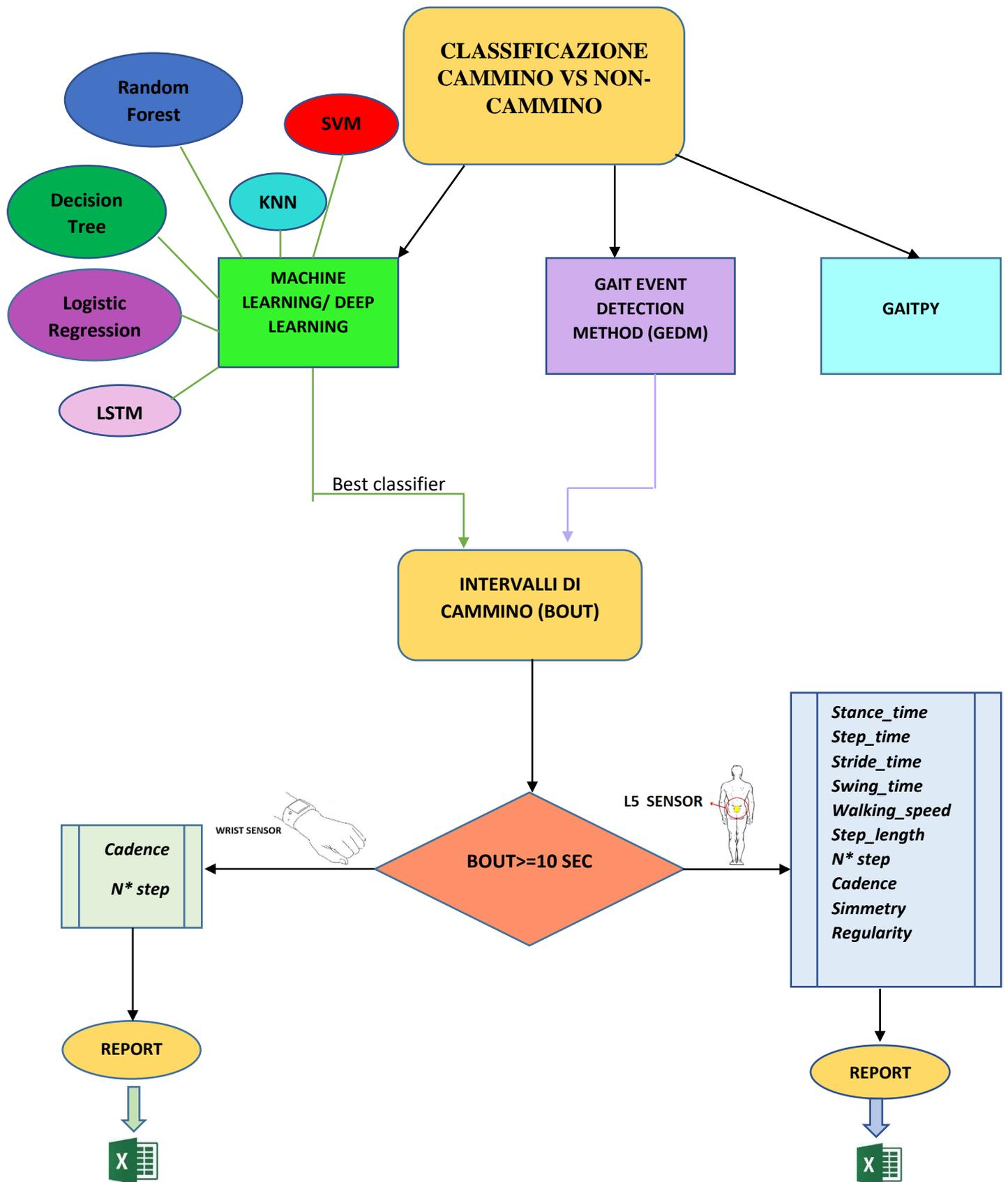


Figura 3.1: Workflow del lavoro svolto.

3.3 Machine Learning applicato al riconoscimento del cammino

Per la classificazione binaria cammino vs non-cammino, step iniziale di questo lavoro di tesi (vedi figura 3.1), sono state adoperate, in primis, tecniche di machine learning.

L'interesse in tali algoritmi è attribuibile a tre loro caratteristiche:

- 1) la loro abilità di generalizzare al modello le complesse relazioni non lineari inerenti dati del cammino;
- 2) la loro predisposizione a lavorare con dati multidimensionali;
- 3) la loro abilità a incorporare facilmente nuovi dati nel cercare di migliorare le performance di predizione.

Il dataset, costituito da dati di accelerazione e velocità angolare registrati sui 16 soggetti partecipanti, è stato suddiviso in due sottogruppi: training-set e test-set in cui si è stata mantenuta una proporzione dei dati del 70% e del 30% rispettivamente.

Nel training-set sono confluite le caratteristiche estratte dai segnali registrati sui primi cinque soggetti mentre il test-set è stato popolato con le feature estratte dai segnali dei rimanenti undici partecipanti.

Le feature utilizzate per gli algoritmi di ML sono state estratte dai dati registrati tramite le IMU posizionate su L5 e sul polso. Per ciascuno dei due sensori sono state estratte un totale di 68 feature a partire dai segnali di accelerazione e di velocità angolare; per la combinazione dei due sensori alle 68 feature se ne sono aggiunte ulteriori 9 per un totale di 145.

In **Tabella 3.2** viene riportata una breve descrizione di tutte le feature che sono state estratte:

Feature	Descrizione feature	Feature	Descrizione feature
1-3	Media dell'accelerazione (x, y, z)	4-6	Varianza dell'accelerazione (x, y, z)
7-9	Correlazione tra gli assi dell'accelerazione (x, y, z)	10-12	Energia di BA (x, y, z)
13	SMA della componente BA	14	Angolo di tilt ottenuto dalla componente GA in direzione verticale
15-17	Media delle componenti di GA (x, y, z)	18	Media di MV della componente BA

19	Varianza di MV della componente BA	20	Energia di MV della componente BA
21-23	Media del segnale di jerk dall'accelerazione (x, y, z)	24-26	Varianza del segnale di jerk dall'accelerazione (x, y, z)
27-29	Correlazione tra gli assi del segnale di jerk dall'accelerazione (x, y, z)	30-32	Energia del segnale di jerk dall'accelerazione (x, y, z)
33	SMA del segnale di jerk dall'accelerazione	34	Media di MV del segnale di jerk dall'accelerazione
35	Varianza di MV del segnale di jerk dall'accelerazione	36	Energia di MV del segnale di jerk dall'accelerazione
37-39	Media della velocità angolare (x, y, z)	40-42	Varianza della velocità angolare (x, y, z)
43-45	Correlazione tra gli assi della velocità angolare (x, y, z)	46-48	Energia della velocità angolare (x, y, z)
49	SMA della velocità angolare	50	Media di MV della velocità angolare
51	Varianza di MV della velocità angolare	52	Energia di MV della velocità angolare
53-55	Media del segnale di jerk dalla velocità angolare (x, y, z)	56-58	Varianza del segnale di jerk dalla velocità angolare (x, y, z)
59-61	Correlazione tra gli assi del segnale di jerk dalla velocità angolare (x, y, z)	62-64	Energia del segnale di jerk dalla velocità angolare (x, y, z)
65	SMA del segnale di jerk dalla velocità angolare	66	Varianza di MV del segnale di jerk dalla velocità angolare
67	Varianza di MV del segnale di jerk dalla velocità angolare	68	Energia di MV del segnale di jerk dalla velocità angolare
69-71	Costante di attenuazione tra le combinazioni dell'accelerazione dei sensori (x, y, z)	72-74	Correlazione tra le combinazioni dell'accelerazione dei sensori (x, y, z)
75-77	Correlazione tra le combinazioni del segnale di velocità angolare dei sensori (x, y, z)		

Tabella 3.2: Feature estratte: feature 1-68 estratte dai dati registrati tramite una singola IMU; feature 69-77 ricavate dalla combinazione dei due sensori.

3.3.1 Feature Selection

La *Feature Selection* (FS) permette di individuare e selezionare le caratteristiche più rilevanti ed informative ai fini della classificazione. Una feature è informativa se è altamente correlata con il target e non correlata con le altre caratteristiche, mentre è considerata rilevante se permette di ottenere buone prestazioni. La feature selection è dunque un'operazione che riduce il set di feature a disposizione consentendo di velocizzare l'esecuzione degli algoritmi, ridurre il costo computazionale e limitare l'utilizzo della memoria, ottenere una migliore comprensione dei processi alla base della generazione dei dati e, in ultima analisi, migliorare le prestazioni del modello limitando l'*overfitting*.

I metodi di feature selection si dividono in tre differenti categorie:

- 1) Metodi *Filter*: selezionano un sottoinsieme di caratteristiche in fase di pre-processing senza tener conto del predittore utilizzato;
- 2) Metodi *Wrapper*: prevedono durante il training l'utilizzo di un subset iniziale di feature a cui vengono aggiunte o rimosse caratteristiche in base ad un criterio di selezione;
- 3) Metodi *Embedded*: l'importanza delle feature è determinata come parte del processo di training.

Nel lavoro sono stati adoperati metodi di feature selection di tipo *Filter*; lavorano direttamente sul set di dati adottando un criterio euristico che dipende unicamente dalle proprietà interne dei dati a disposizione senza considerare lo specifico algoritmo di classificazione. I tre metodi utilizzati fanno parte dell'*ASU feature selection repository* (<http://featureselection.asu.edu>).

3.3.1.1 Correlation Feature Selection (CFS)

È un algoritmo che classifica subset di feature in accordo ad una funzione di valutazione euristica basata sulla correlazione. Un buon subset deve essere costituito da feature altamente correlate con la classe ma non correlate le une con le altre. Avendo bassa correlazione con la classe, le feature irrilevanti vengono ignorate così come avviene per quelle ridondanti le quali vengono depennate a causa della loro alta correlazione con le restanti feature [30].

3.3.1.2 Fast Correlated Based Filter (FCBF)

Questo metodo, adatto a dati di grande dimensionalità, computa la correlazione predominante tra le caratteristiche e le classi e seleziona le feature predominanti eliminando quelle ridondanti [30]. Si basa sul concetto di incertezza simmetrica (SU), definita come il rapporto tra il guadagno di informazione (IG) e l'entropia (H) di due feature x e y :

$$SU(x, y) = \frac{2 \times IG\left(\frac{x}{y}\right)}{(H(x) + H(y))} \quad IG\left(\frac{x}{y}\right) = H(y) + H(x) - H(x, y)$$

Il valore di SU è nel range $[0, 1]$ dove il valore 1 indica che la conoscenza del valore di una feature predice in maniera completa il valore dell'altra mentre un valore pari a 0 indica che X e Y sono indipendenti.

3.3.1.3 ReliefF

Tale metodo ordina le feature in base ai pesi ad esse corrispondenti, assegnati stimando la qualità delle caratteristiche, ovvero la loro capacità di distinguere pattern appartenenti alla medesima classe e pattern appartenenti a classi differenti. A tale scopo, data un'istanza x selezionata in maniera random, Relief cerca le due istanze ad essa più vicine, una della stessa classe detta *nearest hit* (H) e una appartenente alla classe complementare chiamata *nearest miss* (M). Aggiorna quindi la stima del grado di qualità per tutte le feature dipendentemente dai valori di x , M e H .

ReliefF è un'estensione di *Relief* che effettua la stima valutando le prime k istanze più vicine di ogni classe ottenendo il vettore probabilistico dei pesi di ciascuna classe [30].

3.3.2 Sliding Window

La maggior parte dei metodi di classificazione suddivide i segnali adoperati per il calcolo delle feature in più piccoli segmenti temporali. I metodi di suddivisione in periodi più comunemente utilizzati sono due: il *frame-based segmentation* o *sliding window segmentation* e il *period-based segmentation*.

Per tutti gli algoritmi di machine learning è stata adoperata la tecnica di *sliding window*, metodica semplice che non richiede preprocessing dei segnali ma soltanto la suddivisione del segnale in segmenti di lunghezza fissa che prendono il nome di *finestre* [31]. Per ciascuna finestra, vengono derivate le feature, input degli algoritmi di machine learning che associano a ciascuna finestra un'etichetta di classe. Con tale approccio tutti i campioni collocati all'interno di una stessa finestra ricevono la medesima etichetta di classe.

Ai fini della classificazione cammino vs non-cammino, sono state adoperate finestre di 500 campioni, ovvero della lunghezza fissa pari a 5 secondi con overlap del 50%. Come strategia di *window labeling* è stato adottato uno schema *majority vote* in modo tale da assegnare a ciascuna finestra l'etichetta della classe maggiormente presente in essa. Ad esempio, se su 5 secondi di registrazione sono stati rilevati 4 secondi di cammino e 1 secondo di non-cammino, la finestra è stata etichettata come finestra di cammino.

3.3.3 Algoritmi di machine learning testati

Per la classificazione binaria cammino vs non-cammino sono stati impiegati cinque differenti algoritmi di ML che vengono di seguito esplicitati.

3.3.3.1 K-Nearest Neighbour (K-NN)

Lo schema di classificazione prevede la costruzione di uno spazio delle feature multidimensionale nel quale ogni dimensione corrisponde ad una differente feature. Popolato lo spazio con tutti i pattern del training-set, la classificazione di ciascun nuovo pattern avviene identificando i k pattern ad esso più vicini e assegnandogli l'etichetta della classe maggiormente rappresentata nell'intorno individuato. Iperparametro da fornire all'algoritmo è K , ovvero il numero di vicini che partecipano alla votazione; non vi è un valore standard per tale parametro in quanto l'algoritmo è sensibile alla struttura locale dei dati ma è bene che esso assuma un valore dispari soprattutto nei casi di classificazione binaria per evitare di trovarsi in una situazione di parità. In genere, valori più piccoli di k aumentano la varianza e rendono i risultati meno stabili mentre valori più grandi di k aumentano il *bias* ma riducono la sensibilità

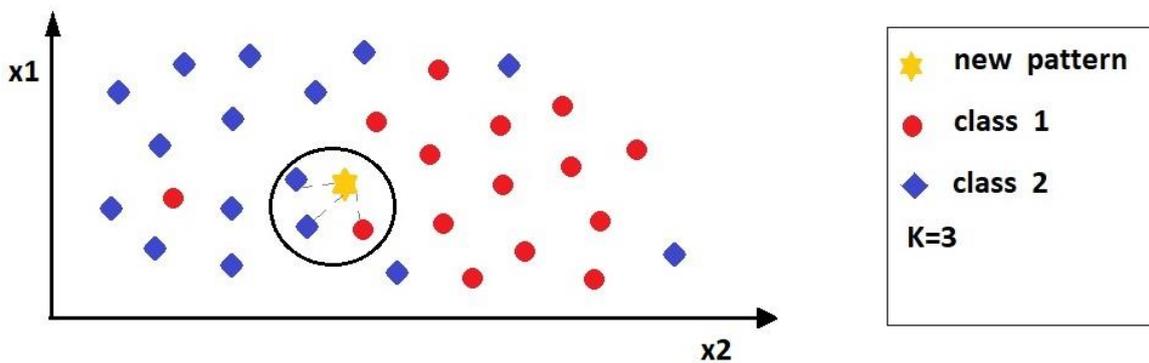


Figura 3.2: Rappresentazione dei pattern del training-set e della nuova istanza da classificare nello spazio delle feature.

Per la classificazione binaria cammino vs non-cammino è stato utilizzato $k=5$ e per il calcolo della distanza della nuova istanza da classificare dai pattern del training-set è stata adoperata la distanza euclidea.

3.3.3.2 Decision Tree

Prevede la creazione di un modello di classificazione tramite la costruzione di un albero decisionale, costituito da una serie di nodi, rami e foglie che viene navigato con approccio *top-bottom* dalla radice alle foglie. Ogni nodo interno denota un test su una feature (o attributo), i rami rappresentano i risultati del test e ogni nodo foglia rappresenta un'etichetta di classe.

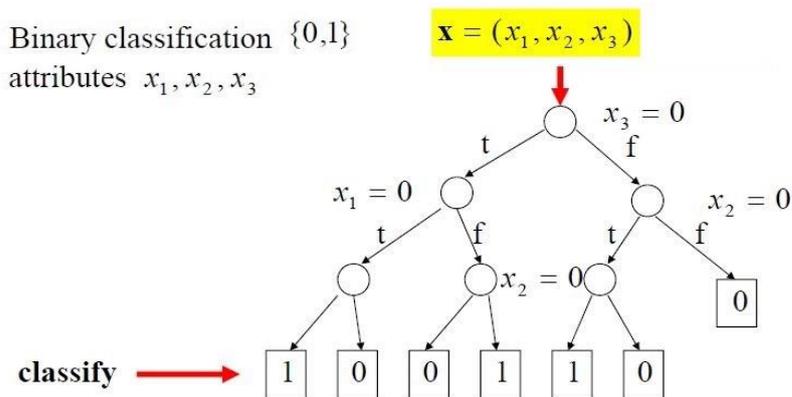


Figura 3.3: Rappresentazione grafica di un albero di decisione per la classificazione binaria.

3.3.3.3 Support Vectore Machine (SVM)

Date due classi di pattern multidimensionali linearmente separabili, SVM identifica, nello spazio delle feature, l'iperpiano in grado di separare le classi con il massimo margine possibile. Quest'ultimo rappresenta la distanza minima dell'iperpiano individuato dai vettori di supporto (pattern più vicini all'iperpiano) e la sua massimizzazione è finalizzata ad avere una migliore generalizzazione sul test-set. Se i pattern nel training-set sono classificati con ampio margine allora si può sperare che anche i pattern del test-set vicini al confine tra le classi siano gestiti in maniera corretta. Quando la separazione lineare non può essere effettuata, allora, adottando una funzione *kernel* è possibile proiettare i dati in uno spazio a più alta dimensionalità dove la separazione ottimale torna ad essere lineare.

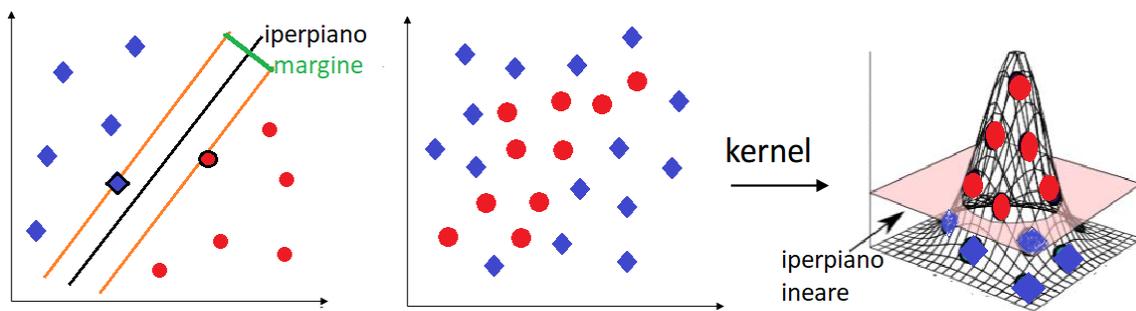


Figura 3.4: Esempio di pattern linearmente separabili nello spazio delle feature (a sinistra); pattern non linearmente separabili (al centro) e pattern linearmente separabili in uno spazio multidimensionale con funzione kernel (a destra).

Per l'esecuzione di questo algoritmo è stata adoperata la libreria LibSVM per Matlab [32], software sviluppato per aiutare gli utenti nell'adoperare SVM in maniera semplice. Per la classificazione è stato scelto il kernel *Radial Basis Function* (RBF) che può essere rappresentato matematicamente come:

$$K(X_1, X_2) = \exp\left(-\frac{\|X_1 - X_2\|^2}{2\sigma^2}\right)$$

dove σ rappresenta la varianza mentre $\|x_1 - x_2\|$ è la distanza euclidea tra due punti. Il valore massimo che può assumere il Kernel è 1 e si verifica quando la distanza tra i due pattern è nulla ovvero i punti coincidono. Quando invece i punti sono separati da una grande distanza, il valore del Kernel è inferiore a 1

e vicino a 0, quindi i punti sono tra loro dissimili. Tale kernel lavora in maniera simile a K-NN ma riesce a superare i problemi della complessità dello spazio poiché deve memorizzare durante l'addestramento unicamente i vettori di supporto e non l'intero set di dati.

3.3.3.4 Random Forest

È un algoritmo appartenente alla famiglia dei metodi di *bagging* che comprende l'aggregazione di più alberi di decisione. Nella sua versione originale, ogni albero è costruito a partire dal dataset originale usando il metodo CART e come criterio di split il Decrease Gini Impurity (DGI) [33]. Nella costruzione di ciascun albero, solo un dato numero di feature selezionate casualmente sono adoperate per lo split. In assenza di tale procedimento, noto come *feature bagging*, molti alberi sceglierebbero con molta probabilità le medesime feature e ciò porterebbe ad avere elevata correlazione tra gli alberi decisionali che costituiscono la foresta. Parametro da impostare per questo algoritmo è il numero di alberi della foresta; sebbene il valore possa variare da alcune centinaia ad alcune migliaia, nella pratica spesso si raggiunge il plateau in termini di performance adoperando qualche migliaia di alberi; nel lavoro è stato adoperato un valore di *ntree* pari a 500.

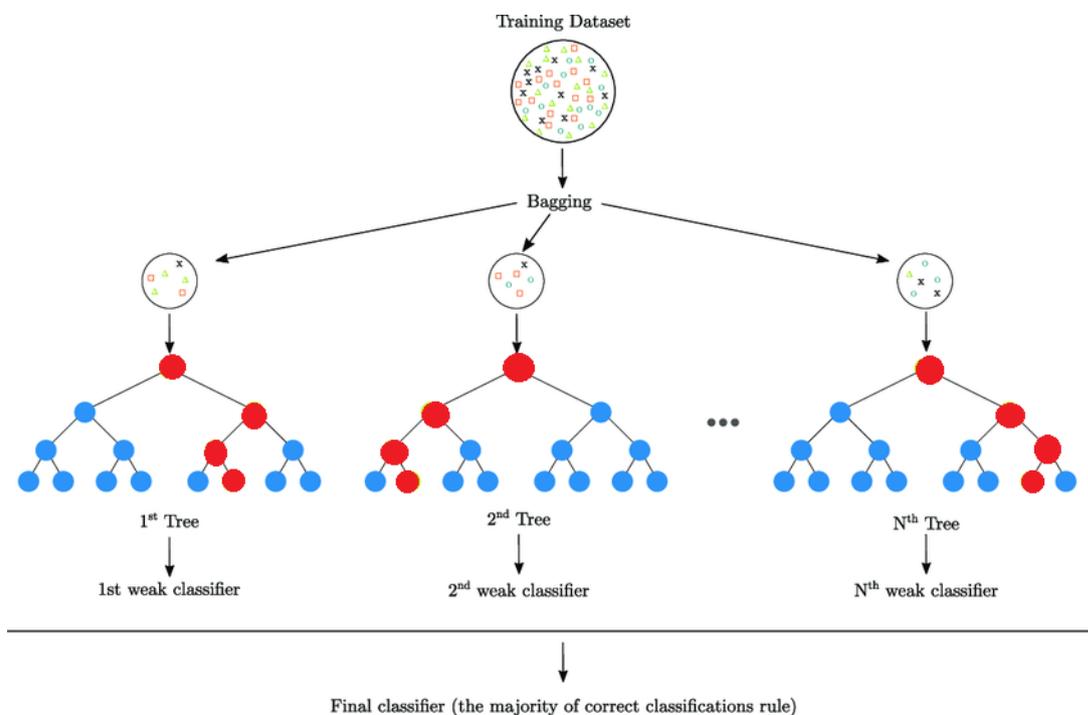


Figura 3.5: Panoramica dell'algoritmo Random Forest.

3.3.3.5 Logistic Regression

Core del metodo è la funzione logistica che può essere adoperata per la classificazione binaria; data un'osservazione x , rappresentata per mezzo di un vettore di feature $[x_1, x_1, \dots, x_n]$ che la caratterizzano, si calcola la probabilità $P(y=1|x)$ che l'osservazione x sia parte della classe $y=1$. La regressione logistica risolve tale compito imparando dal training-set un vettore dei pesi e un termine di bias detto anche intercetta, b . Tali parametri rappresentano i coefficienti dell'algoritmo e vengono calcolati attraverso una stima di massima verosimiglianza. Ogni peso w_i è un numero reale associato ad una feature di input x_i e rappresenta il grado di importanza di quella feature ai fini della decisione di classificazione; può essere positivo (fornendo evidenza che l'istanza che si sta classificando appartiene alla classe positiva) o negativo (fornendo evidenza che l'istanza appartiene alla classe negativa) [33].

Data un'istanza x del test-set si calcola:

$$z = \left(\sum_{i=1}^n w_i \cdot x_i \right) + b$$

La variabile z può assumere valori da $-\infty$ a $+\infty$; per far sì che essa assuma valori compresi in $[0, 1]$ è necessario applicare la funzione sigmoide che mappa i valori di z nell'intervallo desiderato:

$$y = \sigma(z) = \frac{1}{1 + e^{-z}}$$

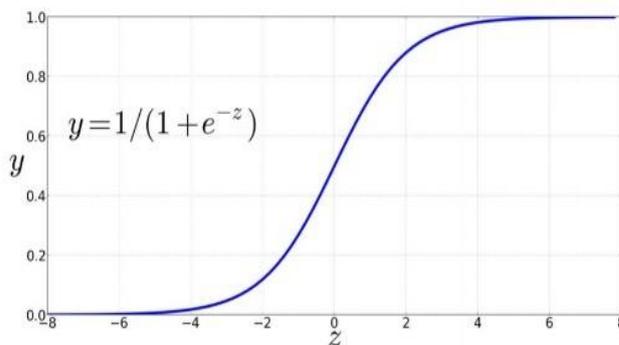


Figura 3.6: Funzione Logistica.

Le probabilità si calcolano come:

$$P(y = 1) = \sigma(w \cdot x + b)$$

$$P(y = 0) = 1 - \sigma(w \cdot x + b)$$

L'algoritmo assegna al pattern x del test-set l'etichetta di classe positiva se la probabilità $P(y=1|x)$ è maggiore di una certa soglia, l'etichetta della classe negativa se il valore di probabilità è minore della soglia. La soglia è tipicamente pari a 0.5, valore utilizzato anche nel progetto.

3.4 Deep Learning per il riconoscimento del cammino

Il Deep Learning è una branca dell'intelligenza artificiale che a partire da una quantità sufficiente di dati è in grado di apprendere la rappresentazione corretta e risolvere problemi di apprendimento automatico grazie all'utilizzo di reti neurali artificiali. A differenza delle tradizionali tecniche di Machine Learning in cui è necessario prevedere uno stadio di pre-processing noto come *feature-extraction* finalizzato alla rilevazione delle feature, gli algoritmi di Deep Learning identificano autonomamente le caratteristiche salienti ai fini della classificazione.

Per i problemi che riguardano la classificazione di attività fisica i metodi di apprendimento profondo possono essere raggruppati in tre categorie: Reti Neurali Convoluzionali (CNN), Reti Neurali Ricorrenti (RNN) e Mixed Model.

3.4.1 LSTM

Per la classificazione binaria è stata scelta la *Long Short Term Memory Network* (LSTM), rete neurale ricorrente che ben si presta alla classificazione di serie temporali. Limite principale delle RNN tradizionali è che le celle di memoria che le costituiscono, dopo un certo lasso di tempo, dimenticano i primi input poiché la loro è una memoria a breve termine (problema del *vanishing gradient*). La rete LSTM è costituita, invece, da celle in grado di considerare le dipendenze a lungo termine tra i dati, quindi, far fronte a tale problematica. Ogni cella possiede 3 tipi di porte: *input gate*, *output gate* e *forget gate* che implementano rispettivamente le funzioni di scrittura, lettura e reset sulla memoria della cella [34].

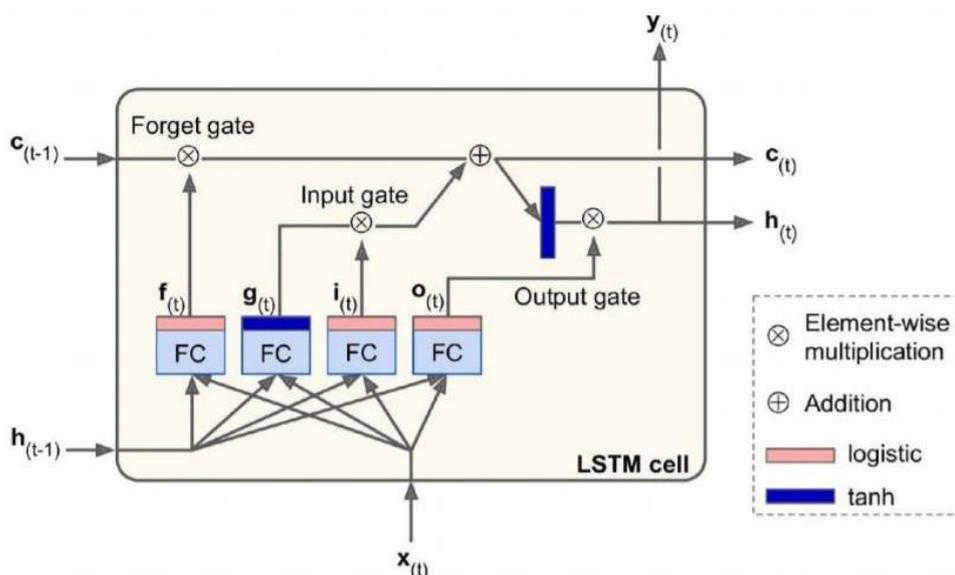


Figura 3.7: Schema a blocchi di una cella LSTM.

La cella funziona in questo modo: il precedente stato (o memoria) a breve termine $h(t-1)$ e i nuovi dati di input vengono inseriti in una rete neurale che prevede una funzione di attivazione sigmoide per la generazione di un vettore in cui ogni elemento è nell'intervallo $[0,1]$. Un valore prossimo a 0 è generato quando la componente di input è considerata irrilevante mentre un valore prossimo a 1 si ha quando l'input è rilevante. I valori emessi sono inviati e moltiplicati in modo puntuale con lo stato (o memoria) a lungo termine passato $c(t-1)$. Il *forget gate* stabilisce quali parti della memoria a lungo termine devono essere dimenticate. Segue poi la fase di decisione che determina quali nuove informazioni devono essere aggiunte alla memoria a lungo termine; la rete neurale è in questo caso attivata dalla funzione \tanh che produce un vettore di aggiornamento della memoria a lungo termine.

L'*input gate* è una rete attivata dalla funzione sigmoide che funge da filtro, identificando quali componenti del nuovo vettore di memoria meritano di essere conservati. Per la generazione dell'output finale, si applica la funzione \tanh allo stato corrente della cella in modo puntuale, si fornisce il precedente stato di memoria a breve termine e i dati di input correnti alla rete neurale attivata da sigmoide per ottenere un vettore filtro che, applicato allo stato della cella mediante moltiplicazione puntuale, fornisce il nuovo stato. La presenza di queste porte permette dunque alla cella di ricordare le informazioni per un tempo indefinito: quando l'input gate è sotto la soglia di attivazione, la cella

mantiene lo stato precedente, mentre se abilitato, lo stato viene combinato con il valore di ingresso. Il forget gate resetta lo stato corrente della cella quando il suo valore viene portato a 0 e l'output gate decide se il valore all'interno della cella deve essere o meno portato in uscita.

Come per gli algoritmi di machine learning, le feature sono state estratte dai segnali accelerometrici e da quelli di velocità angolare. Tuttavia, si è reso necessario effettuare un sottocampionamento dei segnali a 10 Hz in quanto la frequenza di campionamento originale di 100 Hz ha dato luogo ad una mole di dati elevata non supportata che non ha permesso l'esecuzione dell'algoritmo.

3.5 Valutazione delle performance degli algoritmi di classificazione

Dopo aver effettuato il training dei diversi algoritmi e aver testato i modelli sui dati del test-set, sono state calcolate le performance dei diversi classificatori. Gli algoritmi di classificazione vengono comunemente valutati usando rappresentazioni grafiche della performance, ad esempio, la curva ROC o più comunemente per mezzo di una metrica numerica come l'accuratezza. Quest'ultima è ampiamente usata in quanto rappresenta una misura che da sola può riassumere la performance, esprimendo la percentuale di classificazioni corrette. Tuttavia, un'elevata accuratezza non necessariamente caratterizza un buon classificatore soprattutto nei casi di sbilanciamento delle classi. Un'alternativa può essere l'uso delle metriche *Precision* e *Recall*:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{False\ Negative + True\ Positive}$$

Prese singolarmente, nessuna di queste due misure è in grado di catturare in maniera opportuna le performance del classificatore dato che in alcune situazioni si può avere un valore di *precision* alto ma allo stesso tempo un basso valore di *recall* o viceversa. Per valutare le performance dei classificatori è stata, quindi, adoperata la *F-measure* (o *F1-score*), metrica derivata dalla combinazione delle due precedenti [35]:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Per ogni classificatore le performance sono state valutate sul test-set. Avendo a disposizione le etichette vere e quelle fornite in uscita da ciascun classificatore è stato possibile calcolare la matrice di confusione da cui sono state derivate *precision* e *recall*. La media armonica ponderata delle due metriche ha infine fornito il valore di F-measure.

Di seguito vengono mostrati i valori di F-measure ottenuti con i diversi classificatori e i differenti metodi di feature selection.

- ❖ La **Tabella 3.3** riporta le performance ottenute dai diversi classificatori con la soluzione a **singolo sensore su L5**:

<u>L5</u>	K-NN	Decision Tree	Random Forest	SVM	Logistic Regression
Tutte le feature	85,83	82,41	87,44	88,50	87,29
CFS	85,01	80,65	84,84	87,77	85,85
FCBF	73,21	70,74	70,73	80,09	65,57
ReliefF	84,57	80,41	83,60	64,60	81,59

Tabella 3.3: Valori di F-measure espressi in termini percentuali ottenuti per i cinque algoritmi di machine learning con il sensore su L5 adoperando i metodi di feature selection CFS, FCBF e ReliefF.

Dalla tabella si evince che il metodo di feature selection che fornisce i risultati migliori è CFS, per tutti i classificatori; il metodo che fornisce la peggiore performance è FCBF per tutti gli algoritmi ad eccezione di SVM.

La performance migliore, ovvero il valore di F-measure più alto, si ottiene con l'algoritmo di classificazione **SVM** adoperando come metodo di feature selection il Correlation Feature Selection (**CFS**).

- ❖ In **Tabella 3.4** vengono mostrate le performance ottenute dai classificatori con la soluzione a **singolo sensore sul Polso**:

<i>Polso</i>	K-NN	Decision Tree	Random Forest	SVM	Logistic Regression
Tutte le feature	72,60	72,70	76,30	79,97	80,75
CFS	71,93	70,85	73,17	76,53	66,45
FCBF	69,78	66,26	68,47	69,66	61,33
ReliefF	74,24	68,43	72,17	42,11	70,14

Tabella 3.4: Valori di F-measure espressi in termini percentuali ottenuti per i cinque algoritmi di machine learning con il sensore sul Polso adoperando i metodi di feature selection CFS, FCBF e ReliefF.

Anche per il sensore posizionato sul polso, CFS è il metodo di feature selection che fornisce risultati migliori. FCBF fornisce risultati peggiori per tutti i classificatori ad eccezione di SVM dove la performance peggiore si ottiene con il metodo ReliefF. Dai valori riportati in tabella si evince che l'algoritmo di classificazione che performa meglio è **SVM** adoperando come metodo di feature selection **CFS**. Anche Random Forest fornisce un valore di F-measure elevato, di poco dissimile da quello ottenuto con SVM.

❖ In **Tabella 3.5** sono riportati i valori di F-measure ottenuti con la combinazione dei **due sensori, L5+Polso**:

<i>L5+polso</i>	K-NN	Decision Tree	Random Forest	SVM	Logistic Regression
Tutte le feature	82,90	81,72	86,87	86,44	86,28
CFS	82,88	83,68	85,02	86,00	84,75
FCBF	74,56	72,85	74,93	80,91	70,79
ReliefF	82,04	81,74	84,46	45,74	82,42

Tabella 3.5: Valori di F-measure espressi in termini percentuali ottenuti per i cinque algoritmi di machine learning con la combinazione L5+Polso adoperando i metodi di feature selection CFS, FCBF e ReliefF.

Il metodo di feature selection che fornisce risultati migliori continua ad essere CFS, il peggiore è FCBF per K-NN, Decision Tree, Random Forest e Logistic Regression mentre per SVM le performance peggiori si ottengono con ReliefF. I due algoritmi che performano in maniera più efficace sono **SVM e Random Forest**.

❖ È stata inoltre testata una soluzione di tipo **multi-sensore**, le cui performance sono mostrate in **Tabella 3.6** Sono stati presi in esame simultaneamente quattro sensori: alle due IMU precedentemente esaminate (collocate su **L5** e sul **polso**) sono stati aggiunti due sensori inerziali, posizionati rispettivamente sul **petto** e sulla **coscia**. L'analisi è stata effettuata al fine di verificare se l'utilizzo di più sensori fornisse dei risultati, in termini di classificazione, migliori rispetto alla soluzione a singolo sensore e quindi di valutare l'eventuale percentuale di incremento della performance:

<i>L5+polso+petto+coscia</i>	K-NN	Decision Tree	Random Forest	SVM	Logistic Regression
Tutte le feature	87,35	84,94	88,83	88,65	92,14
CFS	86,82	86,48	90,29	88,12	90,48
FCBF	82,30	78,31	80,84	86,28	84,67
ReliefF	88,12	85,57	90,08	46,54	91,38

Tabella 3.6: Valori di F-measure espressi in termini percentuali ottenuti per i cinque algoritmi di machine learning con la soluzione multi-sensore adoperando i metodi di feature selection CFS, FCBF e ReliefF.

Dall'analisi condotta si possono trarre le seguenti conclusioni:

- Tra i due casi di studi a **singolo sensore** esaminati, **il sensore su L5 è la soluzione che fornisce la performance migliore.**
- La soluzione a doppio sensore (L5+Polso) fornisce valori di F-measure più alti rispetto alla soluzione a singolo sensore sul polso ma comparabili con quelli ottenuti con il sensore su L5.
- I risultati ottenuti con la soluzione multi-sensore verificano l'ipotesi di aumentati livelli di performance rispetto alla soluzione a singolo o doppio sensore (vedi Tabella 3.6).

Considerato che la percentuale di incremento che si ha con la soluzione multi-sensore rispetto alla soluzione a singolo sensore su L5 è solo dello 0.4% (vedi Tabella 3.7), appare evidente che utilizzare una soluzione multi-sensore, a scapito di un minore livello di comfort, non è una scelta appropriata al raggiungimento dell'obiettivo di questo lavoro. La scelta dovrebbe, invece,

ricadere sul sensore posizionato su L5 per ottenere una soluzione confortevole in termini di indossabilità e robusta in termini di classificazione.

	SVM L5- multisensore	SVM Polso-multisensore	SVM L5+Polso- multisensore
CFS	0,4%	13%	2,4%

Tabella 3.7: Incremento percentuale delle performance della soluzione multi-sensore rispetto alle soluzioni a singolo sensore e doppio sensore. I risultati si riferiscono all' algoritmo che ha performato meglio in entrambe le soluzioni (SVM con metodo di feature selection CFS).

❖ La **Tabella 3.8** mostra i valori di F-measure ottenuti con la rete neurale profonda LSTM:

L5	POLSO	L5+POLSO
86,13	66,10	83,08

Tabella 3.8: Valori di F-measure espressi in termini percentuali per le soluzioni a singolo e doppio sensore.

- Anche adoperando *l'algoritmo di deep learning* la *performance migliore* tra le soluzioni a *singolo sensore* si ottiene con la IMU *su L5*.
- La soluzione a doppio-sensore fornisce, invece, un valore di F-measure più basso rispetto alla soluzione a singolo sensore su L5 ma più alto rispetto al valore che si ottiene con il singolo sensore posizionato sul polso.

Il peggioramento della performance con una soluzione a due sensori rispetto ad una soluzione a singolo sensore su L5 potrebbe essere dovuto ad un incremento della mole di dati di input non accompagnato da un incremento nel numero dei campioni. Quindi, sarebbe utile verificare se con un incremento del numero dei campioni, tale soluzione darebbe risultati migliori in termini di performance.

Nonostante sia stato eseguito il sottocampionamento dei segnali di accelerazione e velocità angolare non è stato possibile eseguire l'algoritmo LSTM per la soluzione multi-sensore a causa della mole elevata di dati non supportata dal PC.

3.6 Gait Event Detection Method- GEDM

Metodi per il riconoscimento del cammino alternativi al machine learning sono i cosiddetti *Gait Event Detection Method (GEDM)*, approcci che si basano sulla detezione degli eventi associati al cammino [36].

Il GEDM applicato per la classificazione cammino vs non-cammino identifica il cammino a partire dai dati accelerometrici registrati tramite il sensore posizionato su L5 rilevando gli eventi di *heel-strike* (HS), ovvero gli istanti di contatto del tallone con il terreno.

Analogamente a quanto fatto per gli approcci di machine learning, il dataset è stato suddiviso in due subset: test-set formato dai segnali di accelerazione dei soggetti 1-5 e training-set costituito dai dati accelerometrici dei soggetti 6-16. Il training è stato finalizzato alla scelta del valore ottimale del parametro *THa*, ovvero soglia sul segnale accelerometrico che permette l'identificazione dei picchi associati agli eventi di heel-strike.

L'algoritmo lavora sulla norma dell'accelerazione calcolata a partire dalle tre componenti del segnale accelerometrico e successivamente sottocampionata da 100Hz a 40Hz. Al fine di migliorare il rapporto segnale-rumore è stato effettuato un pre-processing in cui è stato eseguito un *detrend* per rimuovere la componente continua ed applicato un filtro FIR passa-basso con frequenza di taglio pari a 3.2 Hz per rendere il segnale più *smooth* e rimuovere contemporaneamente il rumore ad alta frequenza. Poiché è desiderabile avere una risposta in fase nulla in banda passante, l'operazione di filtraggio è stata eseguita due volte (da destra a sinistra e viceversa tramite funzione Matlab *filtfilt*). Per la rilevazione dei picchi del segnale, è stata adoperata una Trasformata Wavelet Continua (CWT) che decompone il segnale in due domini, quello temporale e quello di frequenza. Effettuando un'opportuna scelta della tipologia di CWT e determinando un valore appropriato per il parametro di scala è possibile rimuovere fluttuazioni anomale del segnale preservando al contempo i picchi rappresentativi del fenomeno cercato e il loro *timing*. Il parametro di scala è connesso al concetto di frequenza: per alti valori di scala (ovvero bassi valori di frequenza) le wavelet sono ampie e ciò equivale ad avere informazioni globali del segnale non dettagliate temporalmente; bassi valori di scala (ovvero alti valori di frequenza) corrispondono a wavelet compresse che forniscono visioni dettagliate del segnale nel dominio

temporale. In questo lavoro è stata adoperata una wavelet Gaussiana e un parametro di scala pari a 10 (funzione Matlab *cwt, scale 10, gauss2*).

Un successivo smoothing del segnale è stato ottenuto tramite l'applicazione di un filtro lineare di Savitzky-Golay, filtro passa-basso che sfrutta l'approssimazione polinomiale locale (polinomio di grado 0, lunghezza del frame di smoothing di 3 campioni). I picchi identificati tramite CWT sono stati sottoposti ad un'operazione di *thresholding* adoperando il parametro *THa*: solo i picchi di ampiezza maggiore della soglia sono stati identificati come potenziali eventi di HS. Per i dati del training-set sono stati testati diversi valori di *THa*, da 0.001g a 0.9 g con incremento di 0.01g e per ciascun valore è stato eseguito l'algoritmo di seguito esplicitato per l'identificazione dei periodi di locomozione.

- **Inizializzazione variabili:**

- 1) numero del periodo di locomozione, *LocPer=1*;
- 2) numero di step per ciascun periodo di locomozione, *StepCount_loc_per=0*;
- 3) flag che segnala l'inizio e la fine del periodo di locomozione, *LocFlag=0*;
- 4) durata del periodo di locomozione, *D_step= 0*;
- 5) soglia usata per la comparazione dell'intervallo di tempo che intercorre tra picchi selezionati successivi, *THd=3.5* secondi.

- **Identificazione dei periodi di locomozione**

- Per la prima coppia di potenziali eventi di HS consecutivi, viene valutato il tempo che intercorre tra i due picchi ad essi associati; se tale valore, detto Δt , è minore della soglia *THd*, allora *D_step* viene aggiornata con il valore di Δt . L'inizio del periodo di locomozione *StartLoc_loc_per* viene fatto coincidere con l'istante temporale in cui si verifica il primo picco, *StepCount_loc_per* viene incrementato di un'unità e *LocFlag* viene portato a 1.
- Viene poi nuovamente valutato Δt per la successiva coppia di picchi: se il valore risulta essere ancora minore della soglia *THd*, allora il contatore di step viene incrementato di un'altra unità e la soglia viene aggiornata sommando al valore 1.5 la media di *D_step*.

- L'incremento del numero di step e l'aggiornamento del valore di soglia si ripetono fino a che la condizione sul delta temporale minore della soglia è verificata. Quando tale condizione non è più verificata, vuol dire che il periodo di locomozione è terminato pertanto, la variabile che ne indica la fine, *EndLoc_loc_per*, coincide con l'istante in cui si verifica il primo picco della coppia esaminata.

La soglia *THd* viene nuovamente posta uguale al valore di inizializzazione, il contatore di periodi di locomozione *LocPer* viene incrementato di un'unità e il flag torna ad assumere valore 0.

Di seguito il workflow che mostra i principali step di identificazione dei periodi di locomozione:

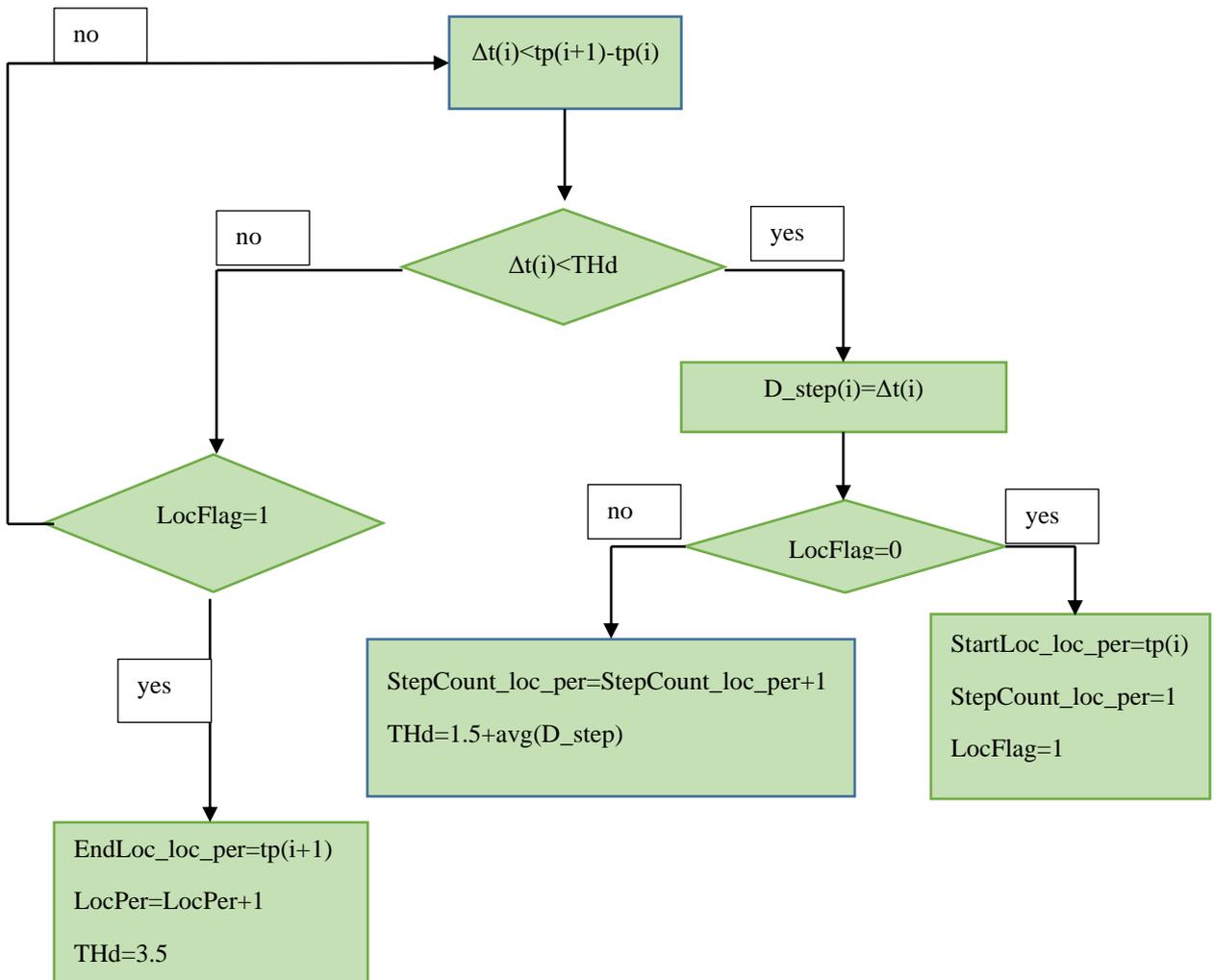


Figura 3.8: Workflow per la rilevazione dei periodi di locomozione.

Ogni intervallo di cammino identificato è caratterizzato da un istante di inizio, un istante di fine e dal numero di step in esso contenuti. Solo i periodi con almeno 4 step consecutivi sono stati ritenuti periodi di reale locomozione e pertanto adoperati nella successiva fase di analisi del cammino.

Per valutare le performance dell'algorithmo è stata attribuita un'etichetta di classe a ciascun campione del segnale accelerometrico. Un'etichetta di valore pari a 1 (ovvero classe *cammino*) è stata attribuita a tutti i campioni collocati all'interno di uno degli intervalli di locomozione identificati mentre ai restanti campioni è stata assegnata un'etichetta di valore pari a 20 (corrispondente alla classe *non-cammino*). Per ciascun valore di THa testato, avendo a disposizione le label vere del dataset e adoperando le etichette attribuite come su descritto, è stata calcolata la matrice di confusione e di conseguenza il valore di F-measure. La valutazione della performance sul training-set ha consentito di scegliere il valore di soglia THa (da utilizzare sui segnali dei soggetti di test) per il quale F1-score è risultato maggiore.

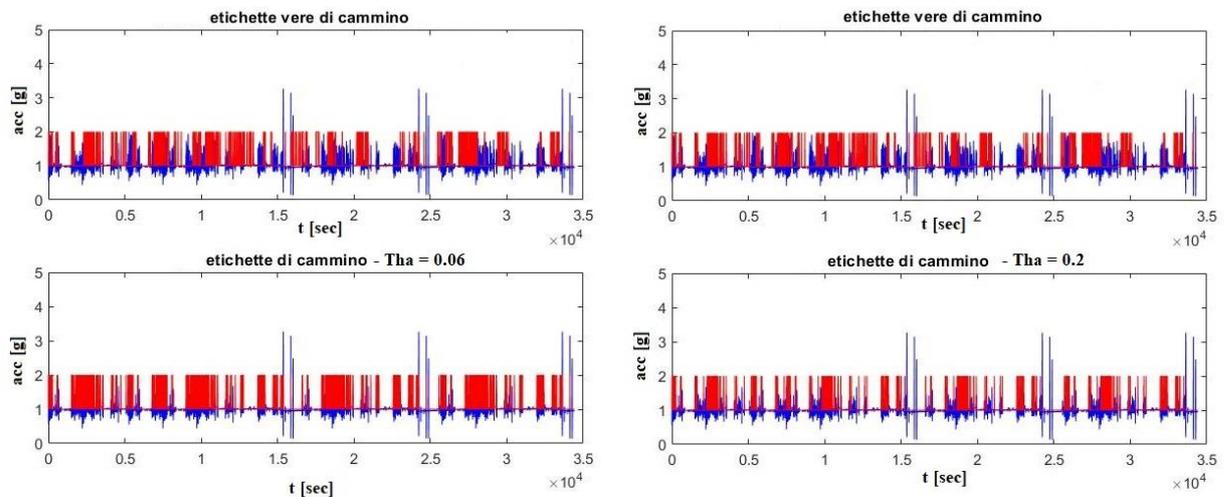


Figura 3.9: Confronto tra le etichette vere del dataset e le label assegnate dall'algorithmo GEDM con due differenti valori di THa .

L'intera procedura di selezione dei picchi, identificazione dei periodi di locomozione e assegnazione delle etichette di classe ai campioni è stata quindi ripetuta sul test-set, ottenendo così un nuovo valore percentuale di F-measure pari a 55,81%.

3.7 Gaitpy

Al fine di confrontare i risultati ottenuti tramite gli approcci di Machine Learning/Deep Learning e GEDM con un metodo di classificazione già sviluppato e validato, è stato adoperato Gaitpy, un pacchetto Python Open Source sviluppato per Pfizer, Inc., compatibile con Python v3.6 su MacOSX, Windows, e Linux e fruibile tramite il programma di installazione Anaconda [37].

Gaitpy implementa un algoritmo con struttura modulare al fine di identificare, a partire dai dati accelerometrici registrati tramite un sensore posizionato in zona lombare, le regioni di cammino ed estrarre per ciascuna di esse feature caratteristiche della locomozione. Le funzioni svolte dall'algoritmo sono tre: classificazione dei periodi di cammino, estrazione di feature cliniche e visualizzazione degli eventi del cammino. Per l'obiettivo fissato è stata utilizzata unicamente la prima delle tre funzioni, tralasciando le restanti due.

A differenza degli approcci visti in precedenza in cui è stato necessario effettuare un training per addestrare i classificatori (per gli approcci di machine learning) o per ricavare il valore ottimale della soglia THa (nel GEDM), per questa soluzione il training non è stato effettuato in quanto l'algoritmo include già in sé un classificatore binario Random Forest pre-allenato.

Gaitpy viene quindi adoperato esclusivamente per i soggetti di test, sui dati accelerometrici registrati dal sensore posizionato su L5.

L'algoritmo ha ricevuto in ingresso un file *csv*, creato tramite Matlab, in cui nella prima colonna è stato memorizzato il *timestamp* e nelle colonne adiacenti le tre componenti dell'accelerazione.

L'esecuzione dell'algoritmo tramite Anaconda ha generato un nuovo file *csv* in cui, per ogni intervallo di cammino identificato con un numero progressivo, ha fornito gli istanti iniziali e finali. Dopo aver importato il file di output in Matlab, analogamente a quanto fatto per l'approccio GEDM, è stata assegnata un'etichetta di valore pari a 1 ai campioni di accelerazione collocati nei periodi di locomozione identificati, etichetta pari a 20 ai campioni rimanenti. Dalle label vere del test-set è stata ricavata la matrice di confusione e quindi calcolato il valore percentuale di F-measure che è risultato pari a 46,80%.

3.8 Confronto delle performance di classificazione

Viene, infine, eseguito un confronto in termini di performance dei diversi approcci adoperati per la classificazione cammino vs non-cammino. Le performance si riferiscono ai risultati ottenuti con il sensore inerziale su L5, essendo questo l'unica collocazione che viene utilizzata da tutte e tre le differenti metodiche (machine learning/deep learning, GEDM, Gaitpy).

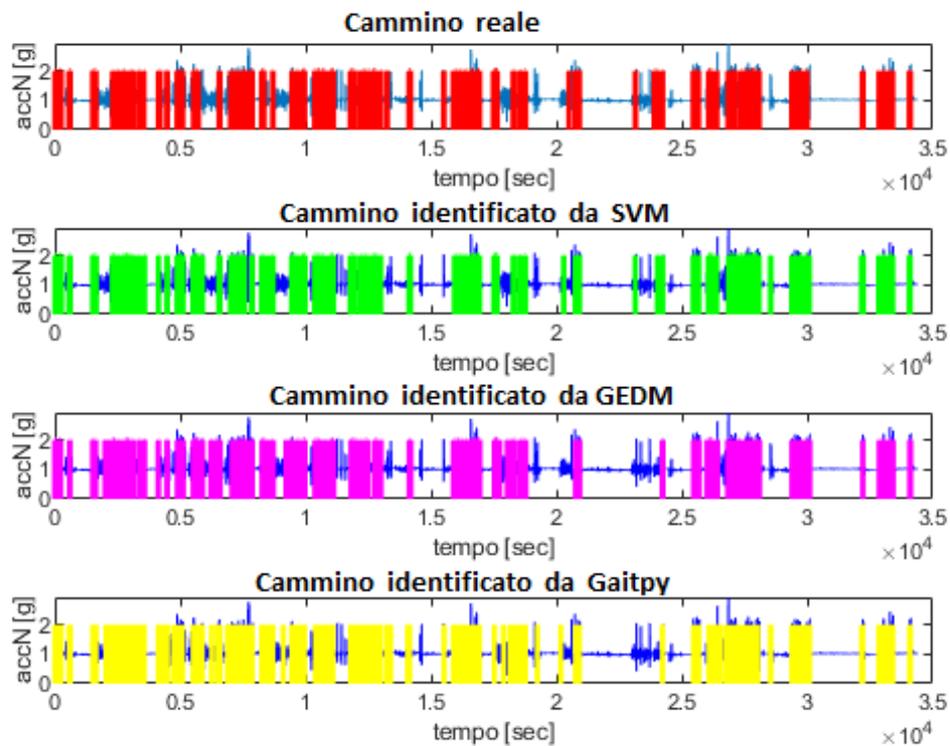


Figura 3.10: Identificazione dei periodi di locomozione con i differenti metodi di riconoscimento del cammino.

- ❖ La **Tabella 3.9** mostra i valori di F-measure ottenuti dai diversi metodi di classificazione; per gli algoritmi di machine learning vengono riportati solo i valori ottenuti adoperando il metodo di feature selection più performante (quello per cui il valore di F-measure ha assunto il valore più alto):

K-NN (cfs)	Decision Tree (cfs)	Random Forest (cfs)	SVM (cfs)	Logistic Regression (reliefF)	LSTM	GEDM	Gaitpy
85,01	80,65	84,84	87,77	87,29	86,13	55,81	46,80

Tabella 3.9: Valori di F-measure espressi in termini percentuali ottenuti dai diversi metodi adoperati per la classificazione, scegliendo per gli algoritmi di machine learning il metodo di feature selection più performante.

- Si può concludere che per la classificazione cammino vs non-cammino gli algoritmi di machine learning e deep learning sono in grado di risolvere il task in maniera più efficiente rispetto all'approccio GEDM e all'Open Source Gaitpy.
- Tutti gli algoritmi di machine learning raggiungono un valore di F-measure superiore all'80% con SVM che si avvicina all'88% mentre i livelli di performance di GEDM e Gaitpy sono inferiori al 60%.

Capitolo 4: Quantificazione e valutazione del cammino

4.1 Dal riconoscimento alla caratterizzazione del cammino

Il processo di Classificazione, che ha rappresentato la prima parte di questo lavoro, è un aspetto di fondamentale importanza che ha permesso di comparare le performance dei diversi classificatori, quindi, individuare gli algoritmi che meglio identificano i periodi di locomozione. Grazie all'identificazione efficace dei periodi di cammino è possibile effettuare un'analisi dettagliata delle caratteristiche del cammino in tali periodi, per ottenere una valutazione oggettiva del cammino che rispecchi la funzionalità del soggetto nella vita di tutti i giorni (*real-world*). Per questo motivo, focus della seconda fase del lavoro (vedi figura 3.1) è rivolto alla stima di parametri in grado di fornire informazioni sulla qualità della locomozione, grandezze che caratterizzano il cammino e che possono contribuire alla creazione di un report oggettivo e facilmente interpretabile da parte del clinico.

4.2 Identificazione degli intervalli (*bout*) di cammino

In prima istanza, sulle regioni precedentemente classificate come cammino/non-cammino, è stata effettuata l'identificazione dei cosiddetti *walking bout (WB)*, intervalli di cammino contraddistinti da un istante iniziale e finale della locomozione, costituiti da segnali accelerometrici da cui possono essere estrapolati parametri per descrivere la qualità del cammino, che possono essere visti come *biomarker* digitali della funzionalità motoria di una persona. Tale procedura è stata portata avanti per l'approccio GEDM e per l'algoritmo di machine learning che ha ottenuto migliori performance durante il processo di classificazione, ovvero quello che ha ottenuto il valore di F-measure più alto. Per il GEDM, l'identificazione dei *bout* è stata eseguita a partire dai dati accelerometrici registrati tramite il sensore su L5, mentre per l'algoritmo di machine learning l'analisi è stata effettuata sia per il sensore su L5 che per quello posizionato sul polso.

La procedura è immediata per la metodologia GEDM: i *bout* sono stati, infatti, identificati in maniera diretta dall'algoritmo stesso il quale ha fornito in uscita

l'istante di inizio e di fine di ciascun periodo di locomozione identificato. Pertanto, effettuando la ricerca di tali istanti all'interno del vettore dei tempi associato ai segnali registrati tramite IMU, si è risaliti agli indici ad essi corrispondenti che hanno permesso di selezionare i campioni del segnale accelerometrico che costituiscono il *bout* di cammino.

Per il metodo di machine learning, invece, la procedura per risalire ai *bout* di cammino è stata più articolata in quanto l'algoritmo ha fornito in uscita non dei campioni di cammino/non-cammino bensì delle finestre di attività. È stato quindi necessario prevedere uno stadio di elaborazione ulteriore per risalire ai campioni del segnale di accelerazione corrispondenti alle finestre. Questo step è stato preceduto da una fase di *re-labeling* delle finestre, strutturato nel modo seguente: se tra due finestre di cammino, output dell'algoritmo di machine learning, è presente un'unica finestra di non-cammino, quest'ultima è considerata finestra di cammino e, allo stesso modo, se tra due finestre di non-cammino è presente un'unica finestra di cammino, questa è etichettata come finestra di non-cammino. In questa fase a ciascuna finestra vengono assegnati 250 indici (che corrispondono ai primi 2,5 secondi di ciascuna finestra), numeri progressivi che possono essere confrontati con quelli associati al vettore dei tempi del segnale accelerometrico. In questo modo, selezionando tutti i campioni appartenenti a finestre di cammino consecutive si ottengono i *bout*.

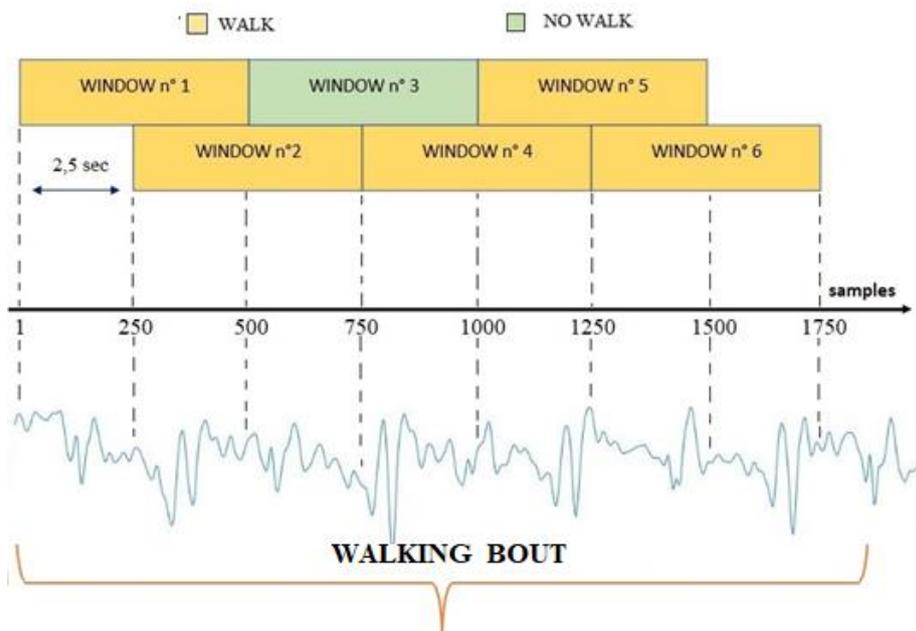


Figura 4.1: Procedura di identificazione di un walking-bout con associazione campioni-finestre di attività.

Tra i bout di cammino identificati, sia per l’algoritmo GEDM che per quello di machine learning, sono stati selezionati solo quelli di durata minima pari a 10 secondi. In linea generale, la durata ottimale è quella che permette di ottenere un numero di cicli del passo sufficiente a consentire una stima rappresentativa e affidabile della performance del cammino. In tal senso, 10 secondi sembrano essere una scelta ragionevole perché rappresentano un intervallo di lunghezza sufficiente a identificare un numero di passi adeguato ad effettuare una stima accurata dei parametri spazio-temporali. Inoltre, uno studio condotto in condizioni di free-living ha mostrato che con bout di durata superiore ai 10 secondi si è in grado di discriminare il cammino di un soggetto affetto da Parkinson da quello di un soggetto sano [38].

In **Figura 4.2** viene mostrata la distribuzione dei *walking bout* nei soggetti partecipanti dove si può osservare che la percentuale di bout di durata inferiore ai 10 secondi è minore rispetto a quella dei bout di durata superiore ai 10 secondi. In questa popolazione, quindi, selezionare bout di 10 secondi, oltre a permettere un’analisi del cammino efficace, consente di analizzare la maggior parte degli intervalli di cammino.

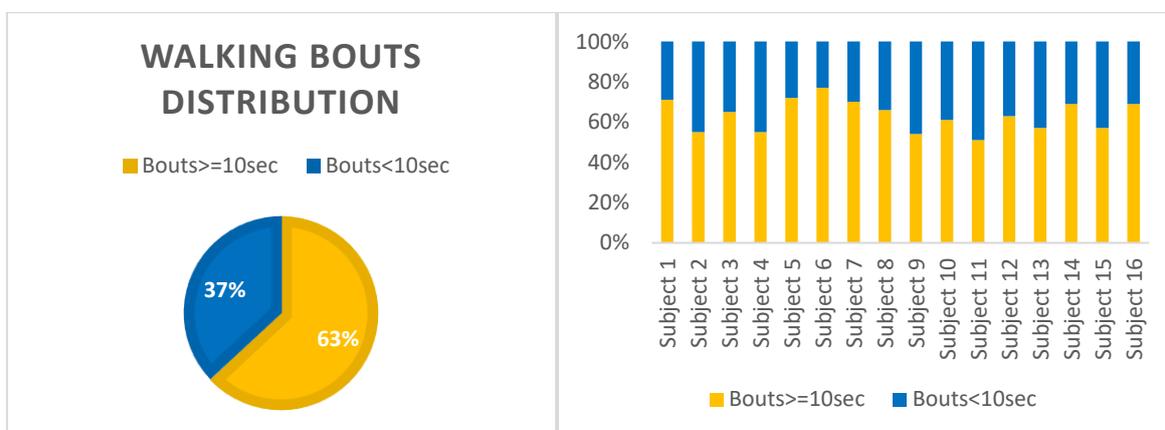


Figura 4.2: Distribuzione dei bout di cammino nei 16 soggetti.

4.3 Parametri che descrivono il cammino

Identificati e selezionati i *bout* di interesse, per ciascuno di essi, il cammino è stato descritto attraverso una serie di parametri che lo caratterizzano. Quantificare, valutare e monitorare tali grandezze è importante dal momento che la loro modificazione può essere associata all’insorgenza di disturbi

neurologici, alla ridotta capacità di mantenimento dell'equilibrio, al rischio di caduta e in linea generale all'abilità funzionale.

L'unità funzionale di riferimento nell'analisi del cammino è rappresentata dal ciclo del passo, definito dall'intervallo di tempo che intercorre tra due contatti iniziali successivi del medesimo piede. Il ciclo del passo si suddivide in due grandi fasi: la fase di *stance* o appoggio e la fase di *swing* o oscillazione. La prima definisce l'intervallo durante il quale il piede rimane a contatto con il terreno e rappresenta circa il 60% del ciclo del passo, mentre la fase di oscillazione definisce l'intervallo durante il quale il piede è sospeso, non impatta con il terreno e rappresenta circa il 40% del ciclo. Per convenzione ogni ciclo del passo inizia con un contatto iniziale (*Initial Contact, IC*) in cui il piede impatta con il terreno tramite l'appoggio del tallone (*Heel Strike, HS*).

I principali parametri spazio-temporali sono:

- 1) *Stride_time* è il tempo che intercorre tra due contatti successivi del tallone (*Heel Strike, HS*) del medesimo piede;
- 2) *Stance_time* è il tempo che intercorre tra il contatto del tallone (*Heel Strike, HS*) e il distacco della punta (*Toe Off, TO*) del medesimo piede;
- 3) *Swing_time* è il tempo che intercorre tra il distacco della punta del piede (*Toe Off, TO*) e il successivo contatto del tallone (*Heel Strike, HS*) del medesimo piede;
- 4) *Step_time* è il tempo che intercorre tra il contatto del tallone di un piede e il contatto del tallone del piede controlaterale;
- 5) *Cadenza* è il numero dei passi al minuto;
- 6) *Step_length* è la distanza percorsa da un singolo arto nel ciclo del passo, ovvero la distanza che vi è tra il punto di appoggio del tallone di un piede e quello del piede controlaterale; è proporzionale a circa il 60% dell'altezza nei soggetti che non presentano alterazioni del cammino;
- 7) *Walking_speed* è il prodotto tra la lunghezza del passo e la cadenza.

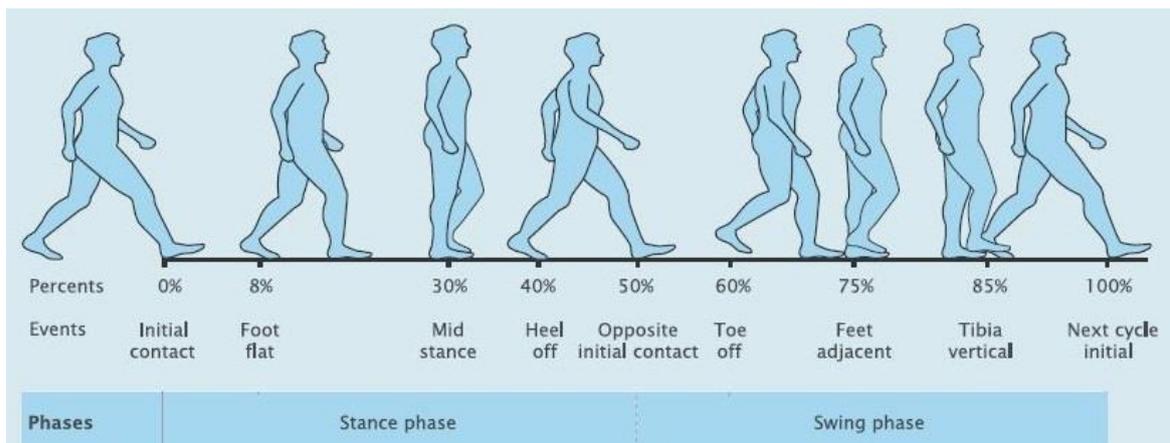


Figura 4.3: Suddivisione del ciclo del passo. Figura adattata da [39].

4.3.1 Stima dei parametri del cammino tramite sensore sulla schiena (L5)

Per la stima dei parametri spazio-temporali del cammino tramite il sensore posizionato su L5 è stato preso in esame il solo segnale accelerometrico; quando un sensore viene posizionato in zona lombare può andare incontro ad un *tilt* a causa dell'intrinseca curvatura di tale regione, dell'eccessiva pronazione del corpo in direzione antero-posteriore nel piano sagittale o ancora a causa di un inaccurato posizionamento del sensore [40]. Il segnale misurato risulta, dunque, essere costituito da due componenti: l'accelerazione statica dovuta all'effetto della gravità (dà una misura dell'inclinazione del sensore rispetto alla verticale) e l'accelerazione dinamica che risulta dal movimento del segmento corporeo al quale la IMU è fissata. Durante il cammino, la media dell'accelerazione dinamica lungo ciascuno dei tre assi è all'incirca nulla quindi, quando la media lungo un asse risulta essere lontana dal valore 0, questa è causata dalla componente statica di gravità lungo quell'asse. Al fine di avere un segnale generato unicamente dal movimento del corpo, è stato necessario effettuare un cambiamento del sistema di coordinate sfruttando la capacità dell'accelerometro di fungere da inclinometro. A tal fine è stato adoperato l'algoritmo sviluppato da Moe-Nilssen che stima le componenti gravitazionali dell'accelerazione nelle direzioni antero-posteriore (AP) e medio-laterale (ML), calcola il *tilt* medio e infine stima le componenti di accelerazione in un sistema di coordinate orizzontale-verticale [40].

La componente antero-posteriore dell'accelerazione a_A nel nuovo sistema di coordinate si può calcolare sfruttando le componenti antero-posteriore a_a e verticale a_v del segnale accelerometrico registrato dalla IMU:

$$\mathbf{a}_A = \mathbf{a}_a \cos \theta_a - \mathbf{a}_v \sin \theta_a$$

Anche la componente medio-laterale a_M può essere stimata a partire dalla corrispondente componente accelerometrica a_m registrata e tramite il vettore a'_v che rappresenta la componente di accelerazione verticale aggiornata tenendo conto della sola accelerazione antero-posteriore nel nuovo sistema di riferimento:

$$\mathbf{a}'_v = \mathbf{a}_a \sin \theta_a + \mathbf{a}_v \cos \theta_a$$

$$\mathbf{a}_M = \mathbf{a}_m \cos \theta_m - \mathbf{a}'_v \sin \theta_m$$

La componente di accelerazione verticale nel nuovo sistema di coordinate orizzontale-verticale è data da:

$$\mathbf{a}_v = \mathbf{a}_m \sin \theta_m + \mathbf{a}'_v \cos \theta_m - \mathbf{g}$$

dove, θ_a e θ_m sono gli angoli tra il piano orizzontale e le accelerazioni misurate in direzione antero-posteriore e medio-laterale, rispettivamente.

Si può dimostrare che $\sin \theta_a$ e $\sin \theta_m$ possono essere approssimati dal valore medio delle accelerazioni in direzione antero-posteriore e medio-laterale.

A seguito delle correzioni effettuate, per le successive elaborazioni, è stata utilizzata solo la componente verticale di accelerazione ottenuta nel nuovo sistema di coordinate. La procedura utilizzata per il calcolo dei parametri spazio-temporali del cammino si basa sulla procedura descritta da Del Din e Godfrey [41].

Per la stima dei parametri temporali del cammino è stato necessario ricavare le sequenze degli eventi di IC (*Initial Contact*) e FC (*Final Contact*) in relazione alla fase di doppio supporto del ciclo del passo.

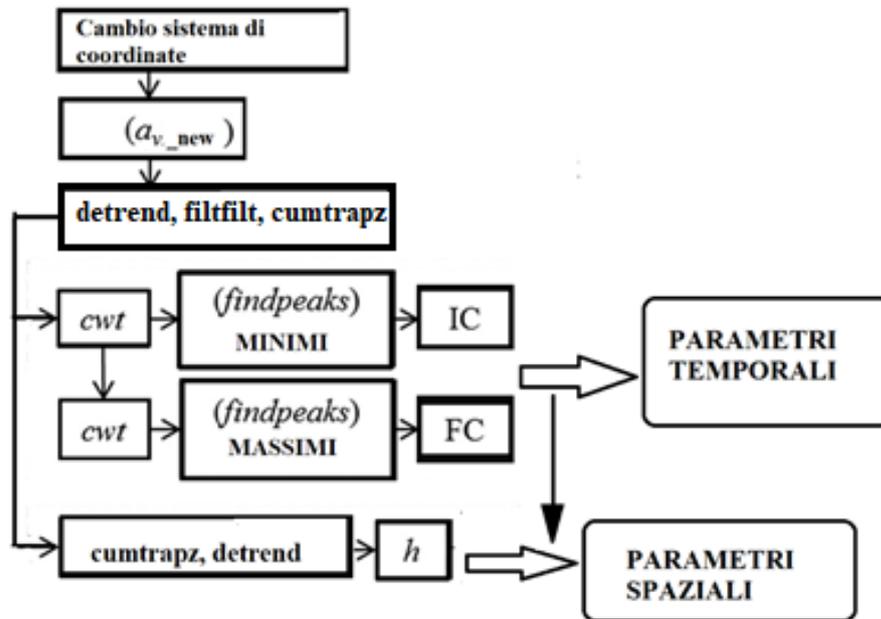


Figura 4.4: Flowchart dell'analisi effettuata per la rilevazione dei parametri spazio-temporali.

Per ridurre il rumore ad alta frequenza associato al segnale è stata eseguita un'operazione di *detrend* sul segnale di accelerazione verticale e successivamente applicato un filtro di Butterworth del 4° ordine a 20 Hz (funzione Matlab *filtfilt*). Il segnale così ottenuto è stato integrato (funzione Matlab *cumtrapz*); l'operazione di integrazione genera inevitabilmente un errore noto come *drift*, segnale non desiderabile a bassa frequenza. Il suo effetto è stato ridotto grazie ad un'operazione di *detrend* e all'applicazione di un filtro di Butterworth passa-alto del 2° ordine.

L'identificazione degli eventi di interesse, ovvero IC ed FC, è stata eseguita tramite analisi *wavelet*. Più nello specifico è stata adoperata una Trasformata Wavelet Continua (CWT), strumento in grado di rilevare contemporaneamente l'evento di interesse e il suo timing. Questa trasformata, inoltre, è in grado prevenire la detezione di eventi spuri che potrebbero verificarsi a causa, ad esempio, del *dragging of feet* [18].

La fase di differenziazione tramite CWT di tipo Gaussiana con parametro di scala pari a 10 (funzione Matlab `cwt, scale 10, gauss2`) ha permesso di identificare gli IC come gli istanti temporali corrispondenti ai minimi del segnale differenziato. Gli FC, invece, sono stati identificati come gli istanti di tempo corrispondenti ai massimi del segnale ulteriormente differenziato tramite CWT [41][42].

Il valore 10 per il parametro di scala è una scelta che viene comunemente effettuata in letteratura, tuttavia, per i segnali esaminati, non si è rivelato efficace nell'identificare correttamente gli istanti di contatto, in modo particolare gli FC. Per tale motivo, la procedura di differenziazione è stata rieseguita incrementando il valore della scala a 16, dando così alla wavelet una forma in grado di identificare in maniera più precisa gli eventi di contatto.

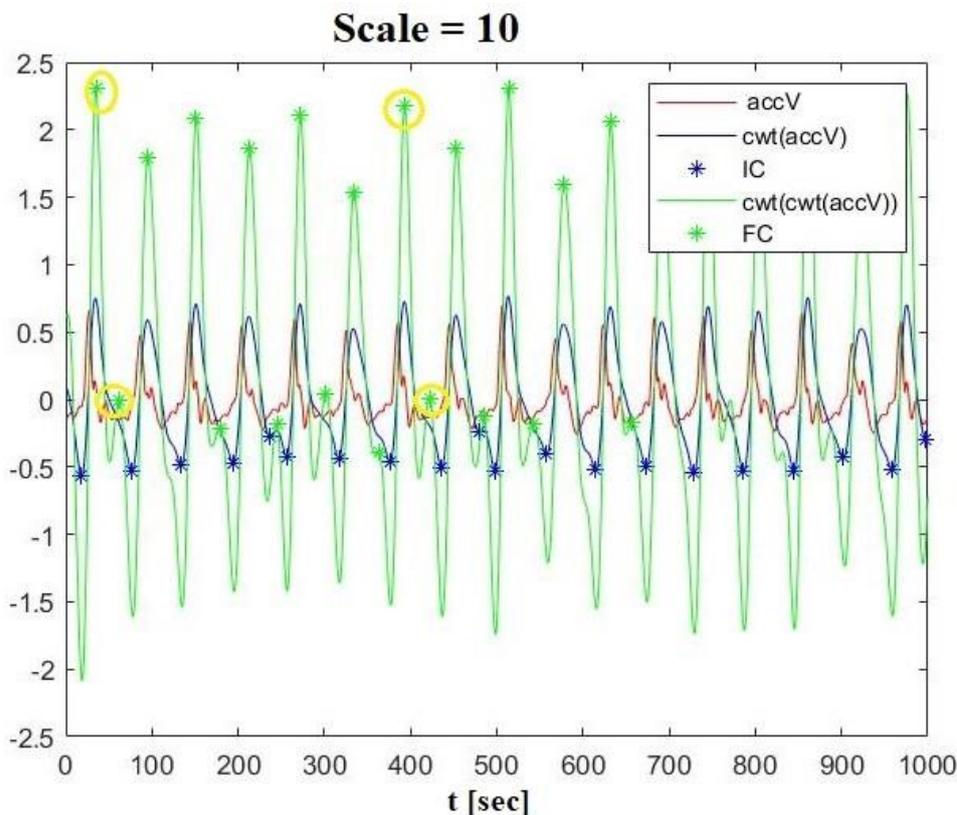


Figura 4.5: Rilevazione degli istanti di IC e FC tramite wavelet con parametro di scala pari a 10; in giallo gli eventi di FC erroneamente identificati.

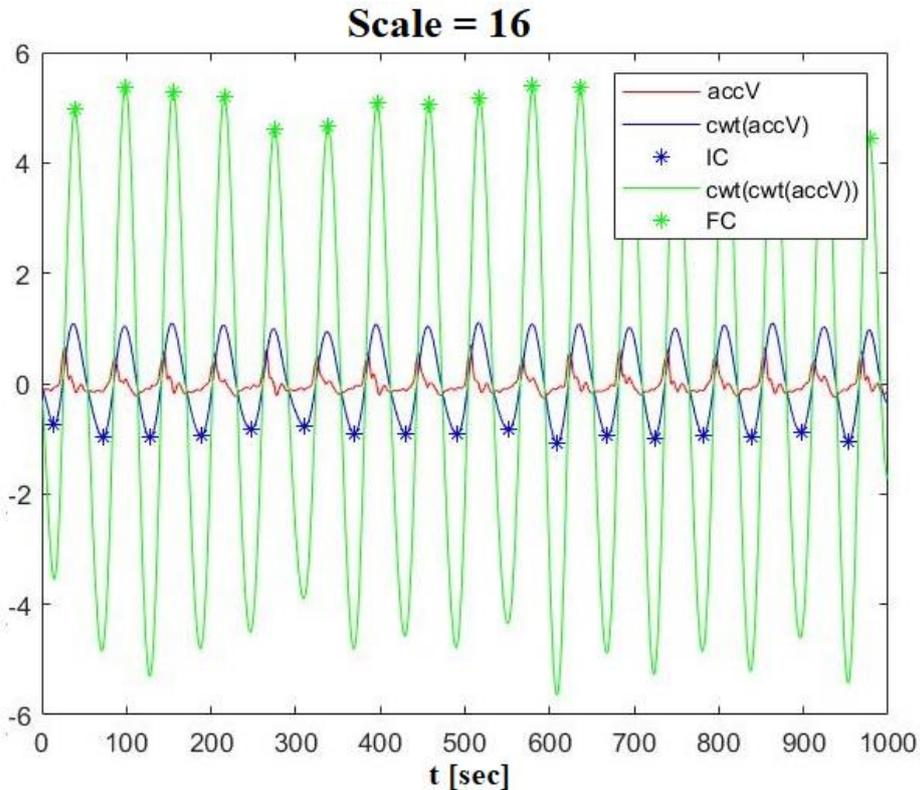


Figura 4.6: Rilevazione degli istanti di IC e FC tramite wavelet con parametro di scala pari a 16.

Ottenute le sequenze di IC ed FC, è stato verificato che il primo istante rilevato corrispondesse ad un IC (istante con cui per convenzione inizia il ciclo del passo). Quando il requisito non è stato soddisfatto, il primo evento rilevato (ovvero un FC) è stato scartato.

A seguito di questa verifica è stata effettuata un'operazione di ottimizzazione finalizzata ad un'identificazione più accurata degli eventi di inizio e fine contatto. Sono stati esclusi tutti gli eventi di IC (e i corrispondenti eventi di FC) che non si sono verificati nell'intervallo temporale 0.25-2.25 secondi dall'evento IC precedentemente rilevato [41].

Eseguita l'ottimizzazione, sono stati calcolati i parametri temporali *Step_time*, *Stance_time*, *Swing_time* e *Stride_time*:

$$Step_time = IC(i+1) - IC(i)$$

$$Stance_time = FC(i+1) - IC(i)$$

$$Swing_time = Stride_time - Stance_time$$

$$Stride_time = IC(i+2) - IC(i)$$

Per la stima dei parametri spaziali del cammino è stato adoperato un algoritmo basato sul modello a pendolo inverso per il tracciamento della traiettoria del Centro di Massa (CoM) del corpo [43]. Tale algoritmo ben predice gli spostamenti del basso tronco durante il cammino. Stimando la variazione in altezza del CoM durante il cammino e nota l'altezza da terra del sensore posizionato in zona lombare, è possibile stimare la *Step_length*:

$$Step_{Length} = 2\sqrt{2lh - h^2}$$

dove l , è l'altezza da terra del sensore e h rappresenta la variazione in altezza del CoM.

Nel processo di acquisizione dati non è stata ricavata l'informazione riguardante la lunghezza del pendolo per cui è stato ipotizzato per il parametro l , un valore pari ad 1 metro. La variabile h è stata stimata attraverso doppia integrazione del segnale di accelerazione verticale. Come in precedenza, anche in questo caso l'integrazione ha dato luogo al *drift* il cui effetto è stato limitato adoperando un filtraggio passa-alto di Butterworth del 4° ordine a 0.1 Hz.

La *Walking_speed* è stata ricavata dal rapporto di *Step_length* e *Step_time*. Questo è un parametro di grande importanza che permette di valutare e monitorare lo stato funzionale e di salute generale di un individuo. Molti modelli predicono il tasso di sopravvivenza valutando un insieme di fattori, tra cui sesso, età, comorbidità, episodi di ospedalizzazione e stato di ipertensione arteriosa; la *Walking_speed*, da sola, si è dimostrata, al pari di tali modelli, capace di predire la sopravvivenza. La sua alta associazione con il tasso di sopravvivenza, in modo particolare nella popolazione anziana, ha fatto sì che fosse definita come il 6° segno vitale [44].

Unitamente ai parametri spazio-temporali del cammino sono state ricavate anche delle grandezze indicative della sua regolarità e simmetria, importanti aspetti nell'analisi del cammino. La simmetria è relazionata alla similarità di *step* (semi-passi) controlaterali consecutivi mentre la regolarità è connessa alla similarità di *stride* (passi) consecutivi [45].

I due parametri sono stati stimati adoperando l'algoritmo proposto da Moen-Nilssen attraverso la funzione di autocorrelazione *unbiased* dell'accelerazione verticale del basso tronco [46].

Un coefficiente di autocorrelazione, $Ad(j)$ è dato dalla somma dei prodotti di una serie temporale $x(i=1,2, \dots, N)$ moltiplicata per una sua replica ritardata nel tempo $x(i+j)$ dove il parametro che indica il ritardo (j) è lo sfasamento espresso in termini di campioni:

$$Ad = \frac{1}{N-j} \sum_{i=1}^{N-j} x(i) \cdot x(i+j)$$

Una stima della funzione di autocorrelazione è data dalla sequenza dei coefficienti di autocorrelazione ottenuti incrementando il ritardo temporale.

Un segnale ciclico produce coefficienti di autocorrelazione con valori di picco in corrispondenza di ritardi temporali uguali al periodo del segnale; tali ritardi sono detti periodi dominanti [46]. Durante la locomozione, i coefficienti di autocorrelazione stimati a partire dal segnale accelerometrico del tronco possono essere utilizzati per quantificare i valori di picco in corrispondenza del primo e del secondo periodo dominante, che rappresentano rispettivamente lo sfasamento relativo allo *step* e allo *stride*.

La funzione di autocorrelazione così calcolata è stata poi normalizzata in modo tale da ottenere una funzione di valore pari a 1 in corrispondenza di ritardo temporale nullo e ne sono stati selezionati i primi due picchi.

Il coefficiente di autocorrelazione in corrispondenza del primo picco dominante ($Ad1$) è un indice di regolarità del segnale di accelerazione tra *step* consecutivi, mentre, il secondo picco ($Ad2$), che si verifica in corrispondenza del secondo periodo dominante, è indice di regolarità tra *stride* consecutivi.

Ricavati i due indici di regolarità, il loro rapporto $Ad1/Ad2$ ha fornito una misura di simmetria del cammino.

4.3.2 Stima dei parametri del cammino con il sensore sul polso

Per tutti i bout di cammino identificati con il sensore posizionato sul polso, sono stati ricavati, a partire dai dati accelerometrici, due parametri: il numero di *step* e la cadenza.

Il segnale accelerometrico corrispondente a ciascun intervallo di cammino è stato filtrato con un filtro di Butterworth del secondo ordine a 10 Hz per limitare l'effetto del rumore ad alta frequenza. Al segnale, successivamente ricampionato a 20 Hz, è stata aggiunta la componente gravitazionale e ne è stata calcolata la norma. Similmente a quanto fatto per i metodi di machine learning, anche per questo algoritmo è stato adoperato un approccio a *sliding window* per la stima dei parametri; nello specifico sono state adoperate finestre di durata pari a 6 secondi (corrispondenti a 120 campioni) con overlap di 5 secondi.

Per ogni finestra è stata calcolata la FFT a 256 punti e adoperata una finestra di Hanning per ridurre il *leakage* in frequenza e ottenere una buona risoluzione in frequenza. Per le successive elaborazioni è stato preso in esame il logaritmo della parte positiva dello spettro e ne è stata effettuata un'interpolazione lineare per aumentare la risoluzione da 0.08Hz a 0.01Hz [47].

Durante il cammino, il movimento delle braccia dovrebbe essere oscillatorio e sincrono con il movimento delle gambe per garantire la migliore funzione di bilanciamento del corpo; il braccio destro dovrebbe muoversi in maniera sincrona con la gamba sinistra e il braccio sinistro con la gamba destra. In questa situazione, nello spettro dell'accelerazione del polso la frequenza fondamentale dovrebbe corrispondere alla frequenza di *stride* mentre la prima armonica dovrebbe rappresentare la frequenza con cui si verificano gli step.

Tuttavia, durante la locomozione, le braccia non necessariamente si muovono in modo regolare assieme alle gambe ma possono anche effettuare movimenti spuri e casuali. Quindi, dipendentemente dal movimento del braccio, la frequenza fondamentale e alcune armoniche possono risultare deboli e sotto il livello del rumore. Dunque, la semplice detezione dei picchi dello spettro non si dimostra una tecnica sufficientemente robusta per stimare la cadenza.

Per tale ragione, è stata adoperata una tecnica in grado di enfatizzare i pattern armonici in frequenza [47] che ha previsto l'uso di filtri Comb e di una funzione di verosimiglianza della cadenza $CL_n[f]$.

Un filtro Comb è un filtro caratterizzato da multiple bande passanti e bande di stop. È molto usato per i segnali vocali dove le bande passanti devono essere centrate a multipli della frequenza fondamentale del parlato, quindi, la risposta in frequenza del filtro deve essere una funzione periodica con periodo uguale

alla frequenza fondamentale. Dato che il linguaggio parlato ha frequenze fondamentali che variano, il filtro deve essere adattabile, ovvero regolato dalla frequenza fondamentale istantanea del segnale vocale [48].

Anche il segnale accelerometrico registrato tramite il sensore sul polso ha un comportamento simile e per enfatizzare i pattern armonici in frequenza sono stati adoperati dei Filtri Comb Adattabili (Adaptive Comb Filter, ACR) così strutturati:

$$H_{ftest}[f] = \begin{cases} 1 & \text{if } f = n \cdot f_{test}, n \in \{1, 2, 3, 4, 5\} \\ -1 & \text{if } f = m \cdot f_{test}, m \in \{1.5, 2.5, 3.5, 4.5\} \\ 0 & \text{else} \end{cases}$$

Dove $f_{test} \in \{0.4, 0.41, 0.42, \dots, 2\} \text{ Hz}$ e $f \in \{0.0, 0.01, 0.02, \dots, 10\} \text{ Hz}$.

Per ogni finestra temporale di 6 secondi, per ogni valore di f_{test} , è stata calcolata $CL_n[f]$ nel modo seguente:

$$CL_n[f_{test}] = \sum_f \text{FFT}_{1000}[f] \cdot H_{ftest}[f]$$

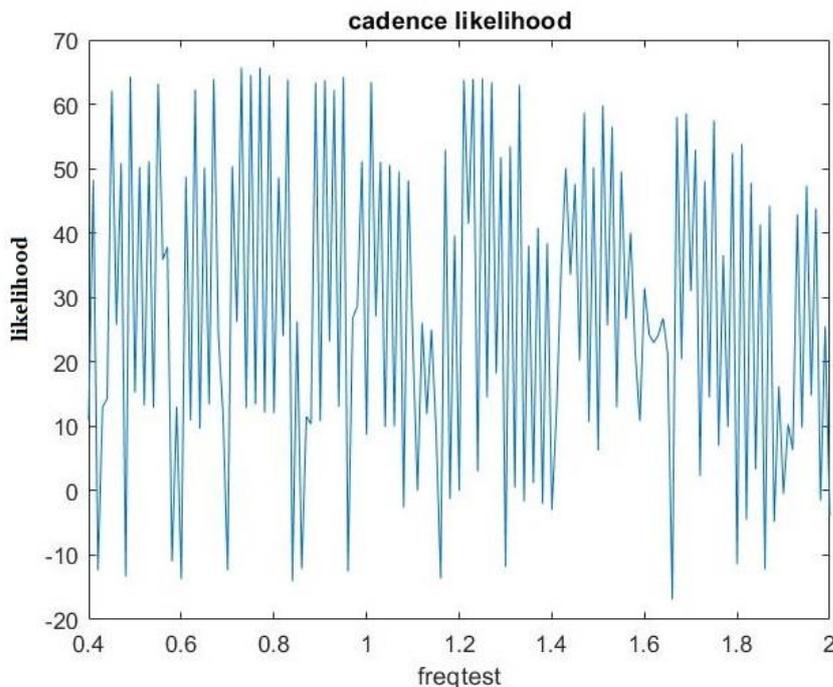


Figura 4.7: Andamento della funzione di Cadence Likelihood.

Come visto in precedenza, in presenza di movimenti del braccio indipendenti dal cammino, $CL_n[f]$ può presentare più di due massimi che non coincidono quindi con la frequenza di *stride* e quella di *step*.

Per evitare la detezione di massimi spuri è stato verificato se la potenziale frequenza di *stride* cambiasse di molto valutando finestre consecutive. In particolare, quando il cambiamento registrato è stato superiore a 0.07 Hz [47], allora il valore di massimo di $CL_n[f]$ è stato ricercato in un intorno della potenziale frequenza di *stride* o di *step*, dipendentemente dal valore di f :

$$C_{wrist}[n] = \begin{cases} 2 \operatorname{argmax}_n CL_n[f] & \text{if } f \in A & \text{with A: } 0.5C_{wrist}[n-1] - 0.01 \leq f \leq 0.5C_{wrist}[n] + 0.07 \\ \operatorname{argmax}_n CL_n[f] & \text{if } f \in B & \text{with B: } C_{wrist}[n-1] - 0.14 \leq f \leq C_{wrist}[n-1] + 0.14 \end{cases}$$

Applicando un filtraggio passa-basso al segnale ricavato e dividendolo per la durata del periodo di cammino, è stata ricavata la cadenza.

Infine, moltiplicando la cadenza per la durata del bout di cammino, è stato ottenuto il valore del parametro di *step*.

4.4 Report

Il lavoro svolto termina con la generazione di un *report* contenente una serie di informazioni oggettive che caratterizzano il cammino precedentemente identificato.

Unitamente all'osservazione clinica diretta e alla professionalità ed esperienza medica, esso può contribuire al processo di *decision making*. Si configura come un valido strumento in grado di incrementare la conoscenza sullo stato di salute geriatrica e di soddisfare il requisito di monitoraggio di lungo periodo in condizioni reali.

Una periodica valutazione dei parametri del cammino e della loro eventuale modificazione in un arco di tempo predefinito può contribuire a rilevare limitazioni funzionali connesse a problematiche dell'apparato locomotore e a diagnosticare precocemente l'insorgenza di malattie neurodegenerative.

Il report prodotto a seguito dell'analisi del cammino consiste di un file *xlsx* generato in maniera automatica dopo aver selezionato il sensore con cui i dati sono stati registrati, scegliendo, quindi, l'unità di misura inerziale posizionata sul polso o quella posizionata in zona lombare. Per entrambi i sensori è possibile selezionare la tipologia di report: un rapporto completo oppure un documento di sintesi.

Sensore sul polso

- Il rapporto completo fornisce informazioni dettagliate riguardanti ciascun *bout* di cammino: numero identificativo progressivo del bout, data e ora di inizio del periodo di locomozione, durata del bout espressa in secondi, numero di step e cadenza.
- Il documento di sintesi, invece, sintetizza le informazioni contenute nel documento completo fornendo media, mediana, deviazione standard e range interquartile di cadenza e numero di step e al contempo riporta la data di acquisizione dei dati e il numero di bout di durata maggiore e inferiore a 10 secondi.

Sensore su L5

- Il rapporto completo fornisce per ciascun bout di cammino: numero identificativo progressivo del bout, data e orario di inizio del periodo di locomozione, numero di step, cadenza, *stance_time*, *stride_time*, *swing_time*, *step_time*, *step_length*, *walking_speed*, parametri di regolarità Ad1 e Ad2 e una misura di simmetria del cammino. Inoltre, può fornire media, mediana, deviazione standard e range interquartile di tutti i parametri spazio-temporali ricavati.
- Il documento di sintesi riassume i dati presenti nel report completo mostrando: data della registrazione, numero di bout di durata inferiore ai 10 secondi, numero di bout di durata maggiore a 10 secondi e indici statistici (media, mediana deviazione standard e range interquartile) di tutti i parametri calcolati.

Walking bouts analysis														
A	B	C	D	E	F	G	H	I	J	K	L	M	N	
bout_number	bout_start	bout_duration_sec	n_steps	cadence	stance_time	stride_time	swing_time	step_time	step_length	speed	Ad1	Ad2	simmetry	
5	11-May-2017 07:03:07	52.5	95	108.571	0.84	1.1	0.26	0.6	0.764212574	1.2737	0.2484	0.6693	0.3711555	
5	11-May-2017 07:03:07	52.5	95	108.571	0.8	1.15	0.35	0.5	0.709813027	1.4196	0.2484	0.6693	0.3711555	
5	11-May-2017 07:03:07	52.5	95	108.571	0.93	1.27	0.34	0.65	0.766070366	1.1786	0.2484	0.6693	0.3711555	
5	11-May-2017 07:03:07	52.5	95	108.571	0.92	1.19	0.27	0.62	0.824520291	1.3299	0.2484	0.6693	0.3711555	
5	11-May-2017 07:03:07	52.5	95	108.571	0.83	1.15	0.32	0.57	0.784126725	1.3757	0.2484	0.6693	0.3711555	
5	11-May-2017 07:03:07	52.5	95	108.571	0.9	1.23	0.33	0.58	0.808701268	1.3943	0.2484	0.6693	0.3711555	
5	11-May-2017 07:03:07	52.5	95	108.571	0.91	1.27	0.36	0.65	0.703020024	1.0816	0.2484	0.6693	0.3711555	
5	11-May-2017 07:03:07	52.5	95	108.571	0.9	1.26	0.36	0.62	0.836317918	1.3489	0.2484	0.6693	0.3711555	
6	11-May-2017 07:04:05	25	43	103.2	0.78	1.13	0.35	0.5	0.279392038	0.5588	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	0.88	1.17	0.29	0.63	0.459490563	0.7294	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	0.81	1.11	0.3	0.54	0.466743437	0.8643	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	1.01	1.38	0.37	0.57	0.842913427	1.4788	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	1.12	1.51	0.39	0.81	0.578896783	0.7147	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	1.03	1.51	0.48	0.7	0.728690495	1.041	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	1.09	1.46	0.37	0.81	0.627969778	0.7753	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	1.03	1.4	0.37	0.65	0.678704795	1.0442	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	1	1.37	0.37	0.75	0.716681377	0.9556	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	1.06	1.57	0.51	0.62	0.747246588	1.2052	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	1.24	1.65	0.41	0.95	0.597393248	0.6288	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	1.01	1.46	0.45	0.7	0.645431324	0.922	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	1.07	1.42	0.35	0.76	0.852030308	1.1211	0.2884	0.2098	1.3747369	
6	11-May-2017 07:04:05	25	43	103.2	0.92	1.15	0.23	0.66	0.727989527	1.103	0.2884	0.2098	1.3747369	

Figura 4.8: Esempio di report completo per il sensore su L5.

Contents

- WALKING REPORT
- PERCENTAGE OF NUMBER OF WALKING BOUTS
- GAIT PARAMETERS
- SPATIO-TEMPORAL GAIT PARAMETERS
- PARAMETERS OF REGULARITY

WALKING REPORT

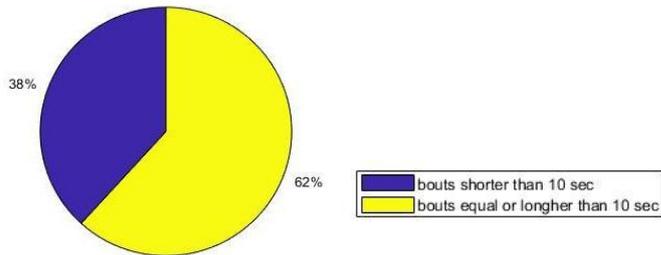
PERCENTAGE OF NUMBER OF WALKING BOUTS

GAIT PARAMETERS

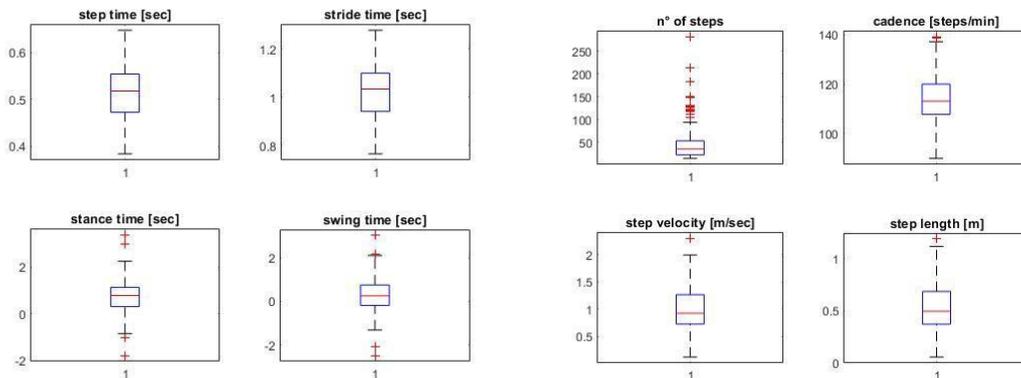
The sensor used for the acquisition of the data is worn on the lower back

The duration of registration is: 573.502 minutes

PERCENTAGE OF NUMBER OF WALKING BOUTS



SPATIO-TEMPORAL GAIT PARAMETERS



PARAMETERS OF REGULARITY



Figura 4.9: Report con boxplot dei parametri spatio-temporali del cammino.

Il clinico, visualizzando e analizzando il report, può mettere a confronto il valore dei parametri ricavati dal monitoraggio del soggetto del paziente con i valori di riferimento per quella fascia d'età al fine di rilevare un *impairment* fisico/cognitivo. Può inoltre tracciare l'evoluzione di tali parametri nel tempo per valutare, ad esempio, se il piano terapeutico implementato è adeguato ed efficace.

4.4.1 Approfondimento sui parametri del cammino

È noto che i parametri spazio-temporali che descrivono il cammino di un soggetto anziano assumono valori diversi da quelli di un soggetto più giovane. Ad esempio, il cammino di una persona anziana è caratterizzato da valori più bassi di *walking_speed*; una riduzione, questa, che può essere collegata al progressivo peggioramento delle abilità funzionali e cognitive che accompagnano l'invecchiamento [49].

La valutazione di tale parametro può aiutare a identificare gli adulti con un aumentato rischio di mortalità precoce. Un cammino lento può riflettere una compromissione del sistema multiorgano (costituito da cuore, polmoni, sistema cardiocircolatorio e muscolo-scheletrico) che per funzionare correttamente richiede un grande dispendio energetico. Tale danno comporta un più alto costo energetico della locomozione determinando una minore propensione a muoversi che, a sua volta, come un circolo vizioso, influenza ulteriormente la salute e pregiudica di conseguenza la sopravvivenza. Sebbene non esista un valore preciso di *cutoff* per discriminare gli anziani che andranno incontro ad un invecchiamento sano o patologico, è stato evidenziato che valori di *walking_speed* inferiori a 0.6 m/s sono responsabili di un aumentato rischio di mortalità mentre valori maggiori di 1 m/s aumentano la probabilità di sperimentare un invecchiamento in salute [7].

Recentemente è stato dimostrato che anche il parametro che esprime il numero di passi al minuto è associato al tasso di mortalità. Alcuni studi hanno messo in luce che un valore di cadenza superiore a 100 passi/min può ridurre il rischio di mortalità del 21% [50].

Una riduzione del parametro di *step_length* può essere riscontrata nella popolazione anziana in soggetti che temono di cadere o hanno già sperimentato un episodio di caduta o ancora in soggetti affetti da osteoartrite o

Parkinson/Alzheimer. Un cammino caratterizzato da ridotti valori di *step_length* è definito anche *cautious gait* [51].

Un ridotto valore di *swing_time* nella popolazione anziana può essere il risultato di una riduzione dell'avanzamento dell'arto durante la fase di swing e di una ridotta potenza muscolare.

Una percentuale più alta della fase di *stance* può essere dovuta ad una ridotta stabilità del tronco e delle estremità inferiori. Un valore più alto di *stance_time* rispetto allo *swing_time* può inoltre indicare che è richiesto un tempo maggiore per mantenere la stabilità del carico a seguito dell'avanzamento dell'arto [52].

Il valore di *step_time* subisce minima variazione durante il cammino; una volta iniziata la marcia i passi si susseguono in maniera continuativa e ritmica.

Un forte incremento del valore di tale grandezza suggerisce una marcia cauta, il timore di cadere o, ancora, può essere sintomo di un disturbo frontale della deambulazione [53].

Similmente a quanto avviene per i parametri spazio-temporali, anche la variazione dei parametri di regolarità e simmetria si pensa possa dipendere dal declino funzionale e neurologico; di conseguenza ci si può aspettare un decremento di tali grandezze nella popolazione anziana. In assenza di modificazioni morfologiche-strutturali del corpo e alterazioni della locomozione, il cammino è pressoché simmetrico. Pertanto, una variazione nei valori nella misura di simmetria può indicare un disturbo neurologico unilaterale o muscoloscheletrico nonché un maggiore rischio di caduta [54].

L'informazione sulla distribuzione della durata dei *walking bout* può risultare utile per verificare se un soggetto è più incline ad eseguire brevi periodi di cammino o se invece propende per intervalli più lunghi.

Se il numero di bout di durata inferiore a 10 secondi rappresenta la percentuale maggiore, è bene monitorarne l'andamento. Un trend di tipo incrementale può evidenziare che, con molta probabilità, il soggetto fa più fatica a muoversi, condizione tipica dei soggetti affetti da Parkinson.

Congiuntamente alle grandezze che descrivono il cammino da un punto di vista quantitativo e qualitativo, anche le informazioni relative alla sua contestualizzazione possono contribuire a definire lo stato di salute di un individuo. Ad esempio, l'informazione riguardante la data della registrazione dei dati può essere adoperata per verificare la presenza di modificazioni

dipendenti dall'attuazione di un piano terapeutico riabilitativo di durata prefissata, dalla stagionalità e da fattori ambientali.

L'orario di inizio dei periodi di locomozione può essere messo in relazione alle diverse fasi della giornata consentendo di individuare se un soggetto è più propenso a camminare di mattina, di pomeriggio o in serata; è possibile, inoltre, verificare se l'assunzione di farmaci influenzi negativamente il cammino o, ancora, se il regime alimentare contribuisca a limitare la propensione a muoversi.

Il *report*, dunque, in ultima analisi, fornisce conoscenza aggiuntiva che favorisce una migliore comprensione da parte del clinico dell'effettivo stato di salute di un soggetto e al contempo consente all'anziano di avere una qualità di vita migliore.

I vantaggi del suo utilizzo possono essere riassunti in:

- 1) monitoraggio di lungo periodo;
- 2) management preventivo;
- 2) riconoscimento precoce degli esordi del declino funzionale o di patologie;
- 3) benessere del soggetto;
- 4) sorveglianza dei fattori di rischio;
- 5) trattamento personalizzato.

Capitolo 5: Conclusioni

5.1 Risultati

Il lavoro di tesi qui presentato si è focalizzato sul riconoscimento del cammino a partire da dati registrati in condizioni di vita reale tramite sensori indossabili posizionati a livello di L5 e del polso. Il processo di classificazione binaria cammino vs non-cammino è stato affrontato con approcci di machine learning/deep learning e con una metodica più strettamente connessa all'elaborazione del segnale.

Per gli algoritmi di machine learning, le feature estratte dai segnali di accelerazione e velocità angolare tramite approccio a sliding window, con diversi metodi di feature extraction, sono state adoperate per eseguire la classificazione sulla base di una soluzione a singolo sensore (L5 o polso), a doppio sensore (L5+polso) o multi-sensore (L5+polso+petto+coscia).

La rete neurale LSTM ha estratto le feature dai dati di accelerazione e velocità angolare in maniera automatica e le ha poi utilizzate per eseguire il riconoscimento del cammino adoperando una soluzione a singolo sensore (L5 o polso) o a doppio sensore (L5+polso).

Il GEDM ha, invece, lavorato unicamente con il segnale di accelerazione registrato tramite il sensore su L5.

Le performance di classificazione delle diverse metodiche, misurate in termini di F-measure, sono state, infine, confrontate con quelle ottenute con l'open-source Gaitpy. La fase di riconoscimento è stata seguita dall'identificazione dei *walking bout* di durata maggiore o uguale a 10 secondi, periodi di locomozione che sono stati sottoposti all'analisi del cammino finalizzata all'estrazione di parametri caratteristici in grado di quantificarlo e descriverne la qualità. Le grandezze estratte dai dati accelerometrici registrati con la IMU posizionata sul polso o su L5 sono state, infine, inserite in un report, file che il clinico può visionare per migliorare la sua conoscenza sullo stato di salute di un soggetto, nonché utilizzare per il processo di *decision-making*.

I risultati mostrano che gli algoritmi di machine learning/deep learning sono metodiche in grado di offrire soluzioni più accurate rispetto ai più tradizionali metodi di riconoscimento basati sulla detezione di eventi del cammino. Inoltre, tutti gli algoritmi testati hanno ottenuto performance migliori anche rispetto ad

un algoritmo già presente e utilizzato in letteratura scientifica (Gaitpy), fornendo così una prova delle buone performance ottenute dagli algoritmi sviluppati.

Tra gli algoritmi di machine learning, SVM e Random Forest si sono dimostrati i più efficaci ai fini della classificazione, fornendo valori di F-measure superiori all'80%; risultati che si sono dimostrati robusti anche nei confronti di suddivisioni random del dataset. Tra le soluzioni a singolo sensore, L5 ha ottenuto il risultato migliore in termini di performance. Anche la soluzione con IMU posizionata a livello del polso ha fornito risultati incoraggianti. La soluzione a doppio sensore (L5+polso) si è dimostrata di poco più performante rispetto alla soluzione a singolo sensore su L5 per gli algoritmi di machine learning ma questo non si è verificato per la rete LSTM. La soluzione multi-sensore non si è rivelata molto più performante della soluzione a doppio e singolo sensore su L5.

Quindi si può concludere che il riconoscimento automatico del cammino attraverso algoritmi di machine learning a partire da dati registrati tramite IMU in condizioni di free-living può rappresentare una valutazione complementare alle tradizionali modalità di acquisizione dei parametri del cammino in laboratorio. Per usufruire al contempo di una soluzione robusta e in grado di garantire il giusto comfort, la soluzione a singolo sensore su L5 dovrebbe essere privilegiata. Anche l'uso di un singolo sensore sul polso dovrebbe essere preso in considerazione nell'ottica di una soluzione che possa offrire ancor più comfort in termini di indossabilità. La scelta di una soluzione multi-sensore non è invece raccomandabile dato il minimo incremento percentuale di performance ottenibile a scapito di una minore indossabilità.

5.2 Discussione

Sebbene i risultati di questo lavoro supportino e promuovano l'uso dei sensori indossabili come strumento affidabile nel riconoscere il cammino ed estrapolarne i suoi parametri, il loro impiego in tale campo dovrebbe essere coadiuvato dall'adozione di dataset di dimensioni maggiori rispetto a quello qui adoperato, al fine di migliorare le capacità di apprendimento da parte degli algoritmi di machine learning e migliorarne quindi le performance di classificazione. Per sostenere lo sviluppo di tali algoritmi sarebbe ideale avere

una grande biobanca di dati digitali sulla mobilità acquisiti su soggetti anziani per lunghi periodi in condizioni di free-living.

Tempi lunghi sono richiesti per l'addestramento di una rete neurale come LSTM; per ridurli sarebbe opportuno utilizzare al posto delle tradizionali CPU le GPU. Unità dotate di migliaia di *core* che possono essere adoperate nel *parallel computing* suddividendo il task da portare a termine in più compiti di dimensione ridotta che possono essere risolti tutti al medesimo tempo.

Sarebbe, inoltre, utile verificare se le performance di classificazione possano migliorare ulteriormente adoperando finestre di dimensione minore/maggiore di quella qui utilizzata per lo sliding window nel processo di feature extraction. Anche la durata dei walking bout è un aspetto che dovrebbe essere ulteriormente indagato al fine di valutare se una maggiore accuratezza nella stima dei parametri sia ottenibile con bout di durata maggiore.

5.3 Prospettive future

Al fine di estendere l'impiego dell'analisi quantitativa del cammino allo screening e alla ricerca clinica, sono necessari tool a basso costo che facilitino le misurazioni in condizioni di free-living. La ricerca futura dovrebbe focalizzarsi sempre più su soluzioni in grado di monitorare il movimento di un soggetto acquisendo dati connessi alla salute in maniera oggettiva e continua durante le attività *day-to-day*, in modo tale da non influenzarne lo svolgimento. In tal senso, l'*ubiquitous computing* offre l'opportunità di avvalersi di accelerometri e giroscopi integrati in device comunemente adoperati nella vita quotidiana, quali smartphone e smartwatch, per mettere a punto un sistema ad hoc per il monitoraggio e la valutazione dello stato di salute di un soggetto. Non bisogna tuttavia dimenticare i limiti connessi all'utilizzo di tali sistemi come, ad esempio, il lento *storage* dei dati, il *battery drain* e le limitate capacità computazionali. Inoltre, particolare attenzione deve essere rivolta alle criticità intrinseche di questa tecnologia: la grande varietà di smartphone in commercio e le differenze in termini di forma e dimensione delle IMU integrate nei dispositivi rendono necessario verificare che le misurazioni diano risultati ripetibili.

Sfida chiave che previene l'inclusione del monitoraggio dell'attività fisica nella pratica/routine clinica è rappresentata dalla necessità di standardizzare i dati tra i differenti device disponibili in commercio, il loro posizionamento e la loro integrazione nel *workflow* clinico.

Il raggiungimento di tale obiettivo darà valore aggiunto a tali sistemi e permetterà al clinico di riporvi maggiore fiducia. Adoperando la tecnologia al servizio della salute, valutando la storia clinica dei soggetti e avvalendosi della conoscenza medica dei professionisti della sanità sarà possibile monitorare, valutare e migliorare lo stato di salute della popolazione anziana perché i processi di prevenzione/diagnosi/terapia risulteranno facilitati.

Appendice

Codice matlab per l'estrazione delle feature

%% FEATURE EXTRACTION A PARTIRE DAI DATI DI ACCELERAZIONE E VELOCITA' ANGOLARE

```
load('usense_back_L5_new.mat') %data recorded with IMU at the lower back
load('shimmer_wrist_down_sampled_new.mat') %data recorded with IMU on the wrist
load('all_sens_data_pos.mat') % Data used to compute across sensor features
load('all_sens_data_pos_gyro.mat'); %Data used to compute across sensor features
load('length_sec')

fs=100; %sample frequency
W=5; %temporal window
wind=W*fs; % window dimension expressed in samples
over_l=0.5*W*fs; % windows overlap
sensors=2;

for S = 1:sensors
    if(S==1)
        activity_sample=shimmer_wrist_down_sampled_new;
    elseif(S==2)
        activity_sample=usense_back_L5_new;
    end
    [m,n] = size(activity_sample);
    if(mean(activity_sample(:,3))<0 && S~=1)
        activity_sample=-activity_sample;
    end
    %% Accelerometer Data Filtering
    order = 3;
    Wp = 0.25 / (fs/2);
    Rp = 0.01;
    Rs = 100;
    [b,a] = ellip(order,Rp,Rs,Wp,'low');
    gravi_signal = filtfilt(b,a,activity_sample);
    BA_signal=activity_sample-gravi_signal;
    %%Jerk computation for Accelerometer and Gyroscope
    %%Accelerometer
    raw_acceleration=activity_sample(:,1);
    Jerk_AccX = compute_jerk_score_from_acceleration(activity_sample(:,1),fs);
    Jerk_AccY = compute_jerk_score_from_acceleration(activity_sample(:,2),fs);
    Jerk_AccZ = compute_jerk_score_from_acceleration(activity_sample(:,3),fs);
    smv_JAcc=Jerk_AccX.^2+Jerk_AccY.^2+Jerk_AccZ.^2;
    %%Gyroscope
    Jerk_GyrX = compute_jerk_score_from_angular_velocity(activity_sample(:,4),fs);
    Jerk_GyrY = compute_jerk_score_from_angular_velocity(activity_sample(:,5),fs);
```

```

Jerk_GyrZ = compute_jerk_score_from_angular_velocity(activity_sample(:,6),fs);
smv_JGyr=Jerk_GyrX.^2+Jerk_GyrY.^2+Jerk_GyrZ.^2;
%SMV Computation Activity recognition using a single accelerometer placed at the wrist or ankle
smv_A_filt=BA_signal(:,1).^2+BA_signal(:,2).^2+BA_signal(:,3).^2;
smv_A=activity_sample(:,1).^2+activity_sample(:,2).^2+activity_sample(:,3).^2;
smv_G=activity_sample(:,4).^2+activity_sample(:,5).^2+activity_sample(:,6).^2;
for i=over_1:over_1:length_sec*fs-over_1
    index=index+1;
    %% Accelerometer Signal
    %Mean Computation from Jerk Signal
    %Accelerometer
    mean_AccX(index)=mean(activity_sample((i-over_1)+1:(i+over_1),1));
    mean_AccY(index)=mean(activity_sample((i-over_1)+1:(i+over_1),2));
    mean_AccZ(index)=mean(activity_sample((i-over_1)+1:(i+over_1),3));
    mean_Acc_smv(index)=mean(smv_A((i-over_1)+1:(i+over_1)));
    %Gyroscope
    mean_GyrX(index)=mean(activity_sample((i-over_1)+1:(i+over_1),4));
    mean_GyrY(index)=mean(activity_sample((i-over_1)+1:(i+over_1),5));
    mean_GyrZ(index)=mean(activity_sample((i-over_1)+1:(i+over_1),6));
    mean_Gyr_smv(index)=mean(smv_G((i-over_1)+1:(i+over_1)));
    % Standard deviation computation from Jerk Signal
    %Accelerometer
    var_AccX(index)=var(activity_sample((i-over_1)+1:(i+over_1),1));
    var_AccY(index)=var(activity_sample((i-over_1)+1:(i+over_1),2));
    var_AccZ(index)=var(activity_sample((i-over_1)+1:(i+over_1),3));
    var_Acc_smv(index)=var(smv_A((i-over_1)+1:(i+over_1)));
    %Gyroscope
    var_GyrX(index)=var(activity_sample((i-over_1)+1:(i+over_1),4));
    var_GyrY(index)=var(activity_sample((i-over_1)+1:(i+over_1),5));
    var_GyrZ(index)=var(activity_sample((i-over_1)+1:(i+over_1),6));
    var_Gyr_smv(index)=var(smv_G((i-over_1)+1:(i+over_1)));
    % Correlation Coefficient
    %Accelerometer
    temp=corrcoef(activity_sample((i-over_1)+1:(i+over_1),1),activity_sample((i-over_1)+1:(i+over_1),2));
    Corr_xyA(index)=temp(1,2);
    temp=corrcoef(activity_sample((i-over_1)+1:(i+over_1),1),activity_sample((i-over_1)+1:(i+over_1),3));
    Corr_xzA(index)=temp(1,2);
    temp=corrcoef(activity_sample((i-over_1)+1:(i+over_1),2),activity_sample((i-over_1)+1:(i+over_1),3));
    Corr_yzA(index)=temp(1,2);
    %Gyroscope
    temp=corrcoef(activity_sample((i-over_1)+1:(i+over_1),4),activity_sample((i-over_1)+1:(i+over_1),5));
    Corr_xyG(index)=temp(1,2);
    temp=corrcoef(activity_sample((i-over_1)+1:(i+over_1),4),activity_sample((i-over_1)+1:(i+over_1),6));
    Corr_xzG(index)=temp(1,2);

```

```

temp=corrcoef(activity_sample((i-over_1)+1:(i+over_1),5),activity_sample((i-
over_1)+1:(i+over_1),6));
Corr_yzG(index)=temp(1,2);
%% Energy of Signal
% Accelerometer
L = wind;
NFFT = 2^nextpow2(L);
energy_Ax(index)=sum(BA_signal((i-over_1)+1:(i+over_1),1).^2);
energy_Ay(index)=sum(BA_signal((i-over_1)+1:(i+over_1),2).^2);
energy_Az(index)=sum(BA_signal((i-over_1)+1:(i+over_1),3).^2);
energy_smvA(index)=sum(smv_A_filt((i-over_1)+1:(i+over_1)).^2);
% Energy of Signal
% Gyroscope
energy_Gx(index)=sum(activity_sample((i-over_1)+1:(i+over_1),4).^2);
energy_Gy(index)=sum(activity_sample((i-over_1)+1:(i+over_1),5).^2);
energy_Gz(index)=sum(activity_sample((i-over_1)+1:(i+over_1),6).^2);
energy_smvG(index)=sum(smv_G((i-over_1)+1:(i+over_1)).^2);
%% Tilt Angle Computation
tilt_ang(index)=real(acos(mean(gravi_signal((i-over_1)+1:(i+over_1),3))));
%% Signal-magnitude area (SMA)
smaA(index)= (sum(abs(BA_signal((i-over_1)+1:(i+over_1),1))+abs(BA_signal((i-
over_1)+1:(i+over_1),2))+...
abs(BA_signal((i-over_1)+1:(i+over_1),3))))/wind;
smaG(index)= (sum(abs(activity_sample((i-
over_1)+1:(i+over_1),4))+abs(activity_sample((i-over_1)+1:(i+over_1),5))+...
abs(activity_sample((i-over_1)+1:(i+over_1),6))))/wind;

%% Gravitational Signal Mean Heike Leutheuser feature
mean_GravX(index)=mean(gravi_signal((i-over_1)+1:(i+over_1),1));
mean_GravY(index)=mean(gravi_signal((i-over_1)+1:(i+over_1),2));
mean_GravZ(index)=mean(gravi_signal((i-over_1)+1:(i+over_1),3));

%% JERK Signal
% Mean Computation from Jerk Signal
% Accelerometer
mean_AccJX(index)=mean(Jerk_AccX((i-over_1)+1:(i+over_1)));
mean_AccJY(index)=mean(Jerk_AccY((i-over_1)+1:(i+over_1)));
mean_AccJZ(index)=mean(Jerk_AccZ((i-over_1)+1:(i+over_1)));
mean_Acc_smvJ(index)=mean(smv_JAcc((i-over_1)+1:(i+over_1)));
% Gyroscope
mean_GyrJX(index)=mean(Jerk_GyrX((i-over_1)+1:(i+over_1)));
mean_GyrJY(index)=mean(Jerk_GyrY((i-over_1)+1:(i+over_1)));
mean_GyrJZ(index)=mean(Jerk_GyrZ((i-over_1)+1:(i+over_1)));
mean_Gyr_smvJ(index)=mean(smv_JGyr((i-over_1)+1:(i+over_1)));
% Standard deviation computation from Jerk Signal
% Accelerometer
var_AccJX(index)=var(Jerk_AccX((i-over_1)+1:(i+over_1)));
var_AccJY(index)=var(Jerk_AccY((i-over_1)+1:(i+over_1)));
var_AccJZ(index)=var(Jerk_AccZ((i-over_1)+1:(i+over_1)));
var_Acc_smvJ(index)=var(smv_JAcc((i-over_1)+1:(i+over_1)));

```

```

% Gyroscope
var_GyrJX(index)=var(Jerk_GyrX((i-over_l)+1:(i+over_l)));
var_GyrJY(index)=var(Jerk_GyrY((i-over_l)+1:(i+over_l)));
var_GyrJZ(index)=var(Jerk_GyrZ((i-over_l)+1:(i+over_l)));
var_Gyr_smvJ(index)=var(smv_JGyr((i-over_l)+1:(i+over_l)));
%% Correlation Coefficient Article
% Accelerometer
temp=corrcoef(Jerk_AccX((i-over_l)+1:(i+over_l)),Jerk_AccY((i-over_l)+1:(i+over_l)));
Corr_xyJA(index)=temp(1,2);
temp=corrcoef(Jerk_AccX((i-over_l)+1:(i+over_l)),Jerk_AccZ((i-over_l)+1:(i+over_l)));
Corr_xzJA(index)=temp(1,2);
temp=corrcoef(Jerk_AccY((i-over_l)+1:(i+over_l)),Jerk_AccZ((i-over_l)+1:(i+over_l)));
Corr_yzJA(index)=temp(1,2);
% Gyroscope
temp=corrcoef(Jerk_GyrX((i-over_l)+1:(i+over_l)),Jerk_GyrY((i-over_l)+1:(i+over_l)));
Corr_xyJG(index)=temp(1,2);
temp=corrcoef(Jerk_GyrY((i-over_l)+1:(i+over_l)),Jerk_GyrZ((i-over_l)+1:(i+over_l)));
Corr_yzJG(index)=temp(1,2);
temp=corrcoef(Jerk_GyrX((i-over_l)+1:(i+over_l)),Jerk_GyrZ((i-over_l)+1:(i+over_l)));
Corr_xzJG(index)=temp(1,2);
%% Energy of Jerk Signal
energy_JAx(index)=sum(Jerk_AccX((i-over_l)+1:(i+over_l)).^2);
energy_JAy(index)=sum(Jerk_AccY((i-over_l)+1:(i+over_l)).^2);
energy_JAz(index)=sum(Jerk_AccZ((i-over_l)+1:(i+over_l)).^2);
energy_smvJA(index)=sum(smv_JAcc((i-over_l)+1:(i+over_l)).^2);
energy_JGx(index)=sum(Jerk_GyrX((i-over_l)+1:(i+over_l)).^2);
energy_JGy(index)=sum(Jerk_GyrY((i-over_l)+1:(i+over_l)).^2);
energy_JGz(index)=sum(Jerk_GyrZ((i-over_l)+1:(i+over_l)).^2);
energy_smvJG(index)=sum(smv_JGyr((i-over_l)+1:(i+over_l)).^2);
%% SMA of Jerk Signal
smaJA(index)=(sum(abs(Jerk_AccX((i-over_l)+1:(i+over_l)))+abs(Jerk_AccY((i-over_l)+1:(i+over_l)))+abs(Jerk_AccZ((i-over_l)+1:(i+over_l)))))/wind;
smaJG(index)=(sum(abs(Jerk_GyrX((i-over_l)+1:(i+over_l)))+abs(Jerk_GyrY((i-over_l)+1:(i+over_l)))+abs(Jerk_GyrZ((i-over_l)+1:(i+over_l)))))/wind;
if(S==1)
    %% Features computation across sensors
    % Mediolateral
    data=all_sens_data;
    cWLM(index)=(1-rms(data((i-over_l)+1:(i+over_l),10))/rms(data((i-over_l)+1:(i+over_l),1)))*100;
    % Anteroposterior
    cWLa(index)=(1-rms(data((i-over_l)+1:(i+over_l),11))/rms(data((i-over_l)+1:(i+over_l),2)))*100;
    % Vertical
    cWLv(index)=(1-rms(data((i-over_l)+1:(i+over_l),12))/rms(data((i-over_l)+1:(i+over_l),3)))*100;
    %% Correlation computation across sensors Accelerometer
    % Mediolateral
    temp=corrcoef(data((i-over_l)+1:(i+over_l),1),data((i-over_l)+1:(i+over_l),10));
    CorrWLM(index)=temp(1,2);% (BW Wrist and L5)

```

```

    % Anterioposterior
    temp=corrcoef(data((i-over_l)+1:(i+over_l),2),data((i-over_l)+1:(i+over_l),11));
    CorrWLa(index)=temp(1,2);%(BW Wrist and L5)
    % Vertical
    temp=corrcoef(data((i-over_l)+1:(i+over_l),3),data((i-over_l)+1:(i+over_l),12));
    CorrWLv(index)=temp(1,2);%(BW Wrist and L5)
    %% Correlation computation across sensors Gyroscope
    dataG=all_sens_dataG;
    % Mediolateral
    temp=corrcoef(dataG((i-over_l)+1:(i+over_l),1),dataG((i-...
over_l)+1:(i+over_l),10));
    GCorrWlm(index)=temp(1,2);%(BW Wrist and L5)
    % Anterioposterior
    temp=corrcoef(dataG((i-over_l)+1:(i+over_l),2),dataG((i-...
over_l)+1:(i+over_l),11));
    GCorrWLa(index)=temp(1,2);%(BW Wrist and L5)
    % Vertical
    temp=corrcoef(dataG((i-over_l)+1:(i+over_l),3),dataG((i-...
over_l)+1:(i+over_l),12));
    GCorrWLv(index)=temp(1,2);%(BW Wrist and L5)
end
end
end

```

%% Funzione per il calcolo del jerk dai segnali di accelerazione

```

function jerk = compute_jerk_score_from_acceleration(acc,fs)
    [bn,an] = butter(4,20/(fs /2));
    acceleration = filtfilt(bn,an,acc);
    temp = acceleration(3:length(acceleration)) - acceleration(1:length(acceleration)-2);
    jerk = (fs/2)*[acceleration(2)-acceleration(1); temp./2 ;acceleration(length(acceleration))-
acceleration(length(acceleration)-1)];
end

```

%% Funzione per il calcolo del jerk dai segnali di velocità angolare

```

function jerk2 = compute_jerk_score_from_angular_velocity(angular_velocity,fs)
    [bn,an] = butter(4,20/(fs /2));
    angular_velocity = filtfilt(bn,an,angular_velocity);
    temp = angular_velocity(3:length(angular_velocity)) - angular_velocity(1:length(angular_velocity)-2);
    jerk = (fs/2)*[angular_velocity(2)-angular_velocity(1); temp./2
;angular_velocity(length(angular_velocity))-angular_velocity(length(angular_velocity)-1)];
    temp2 = jerk(3:length(jerk)) - jerk(1:length(jerk)-2);
    jerk2 = fs*[jerk(2)-jerk(1); temp2./2 ;jerk(end)-jerk(end-1)];
end

```

Bibliografia

- [1] W. H. Day, “World Health Day 2012 focuses on ageing and health,” *East. Mediterr. Health J.*, vol. 18, no. 4, p. 303, 2012.
- [2] United Nations Department of Economic and Social Affairs Population Division, *World Population Ageing 2019*, vol. Highlights. 2019.
- [3] UNECE, “Challenges and opportunities of population ageing,” 2007, [Online]. Available: https://www.unece.org/fileadmin/DAM/highlights/what_ECE_does/English/0726054_UNECE_AGEING.pdf.
- [4] S. Ganguly, “Good health adds life to years,” *J. Indian Med. Assoc.*, vol. 110, no. 4, pp. 212–213, 2012.
- [5] P. Nell and S. S. Nazionale, *Ministero della Salute Ministero della Giustizia Linee di indirizzo sull’attività fisica per le differenti fasce d’età e con riferimento a situazioni fisiologiche e fisiopatologiche e a sottogruppi specifici di popolazione*. 2008.
- [6] J. D. Omura, E. N. Ussery, F. Loustalot, J. E. Fulton, and S. A. Carlson, “Walking as an opportunity for Cardiovascular disease prevention,” *Prev. Chronic Dis.*, vol. 16, no. 5, pp. 1–9, 2019.
- [7] G. V. Ostir *et al.*, “Gait Speed and Dismobility in Older Adults,” *Arch. Phys. Med. Rehabil.*, vol. 96, no. 9, pp. 1641–1645, 2015, doi: 10.1016/j.apmr.2015.05.017.
- [8] A. Piau, N. Mattek, R. Crissey, Z. Beattie, H. Dodge, and J. Kaye, “When Will My Patient Fall? Sensor-Based In-Home Walking Speed Identifies Future Falls in Older Adults,” *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.*, vol. 75, no. 5, pp. 968–973, 2020, doi: 10.1093/gerona/glz128.
- [9] R. McArdle, R. Morris, J. Wilson, B. Galna, A. J. Thomas, and L. Rochester, “What Can Quantitative Gait Analysis Tell Us about Dementia and Its Subtypes? A Structured Review,” *J. Alzheimer’s Dis.*, vol. 60, no. 4, pp. 1295–1312, 2017, doi: 10.3233/JAD-170541.
- [10] “How you walk could be used to identify some types of dementia.” .

- [11] “Gait analysis as a clinical tool for dementia_ current perspectives and future challenges - ACNR _ Paper & Online Neurology Journal ACNR _ Paper & Online Neurology Journal.” .
- [12] R. Codivilla, “Conferenza Nazionale di Consenso Appropriatelyzza clinica e metodologica dell ’ analisi strumentale del cammino (‘ Gait Analysis ’) con particolare Istituto Ortopedico Rizzoli,” pp. 1–37, 2013.
- [13] A. Godfrey, S. Del Din, G. Barry, J. C. Mathers, and L. Rochester, “Instrumenting gait with an accelerometer: A system and algorithm examination,” *Med. Eng. Phys.*, vol. 37, no. 4, pp. 400–407, 2015, doi: 10.1016/j.medengphy.2015.02.003.
- [14] D. Kobsar *et al.*, “Validity and reliability of wearable inertial sensors in healthy adult walking: A systematic review and meta-analysis,” *J. Neuroeng. Rehabil.*, vol. 17, no. 1, pp. 1–21, 2020, doi: 10.1186/s12984-020-00685-3.
- [15] U. Della Croce, A. Peruzzi, and A. Cereatti, “La valutazione e l ’ intervento riabilitativo del cammino mediante l ’ uso di sensori a basso costo Il ciclo della riabilitazione,” no. March 2017, pp. 26–29, 2016.
- [16] A. K. Bourke, E. A. F. Ihlen, R. Bergquist, P. B. Wik, B. Vereijken, and J. L. Helbostad, “A physical activity reference data-set recorded from older adults using body-worn inertial sensors and video technology— The ADAPT study data-set,” *Sensors (Switzerland)*, vol. 17, no. 3, 2017, doi: 10.3390/s17030559.
- [17] M. Awais, L. Chiari, E. A. F. Ihlen, J. L. Helbostad, and L. Palmerini, “Physical Activity Classification for Elderly People in Free-Living Conditions,” *IEEE J. Biomed. Heal. Informatics*, vol. 23, no. 1, pp. 197–207, 2019, doi: 10.1109/JBHI.2018.2820179.
- [18] H. A., D. D. S., R. L., and G. A., “Detecting free-living steps and walking bouts: validating an algorithm for macro gait analysis,” *Physiol. Meas.*, vol. 38, no. 1, pp. N1–N15, 2017, [Online]. Available: <http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L619757013%0Ahttp://dx.doi.org/10.1088/1361-6579/38/1/N1>.
- [19] K. B. Friesen, Z. Zhang, P. G. Monaghan, G. D. Oliver, and J. A. Roper, “All eyes on you: how researcher presence changes the way you walk,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–8, 2020, doi: 10.1038/s41598-020-73734-5.
- [20] A. Saboor *et al.*, “Latest research trends in gait analysis using wearable

- sensors and machine learning: A systematic review,” *IEEE Access*, vol. 8, pp. 167830–167864, 2020, doi: 10.1109/ACCESS.2020.3022818.
- [21] N. F. Ghazali, N. Shahar, N. A. Rahmad, N. A. J. Sufri, M. A. As’ari, and H. F. M. Latif, “Common sport activity recognition using inertial sensor,” *Proc. - 2018 IEEE 14th Int. Colloq. Signal Process. its Appl. CSPA 2018*, no. March, pp. 67–71, 2018, doi: 10.1109/CSPA.2018.8368687.
- [22] T. Zhen, L. Mao, J. Wang, and Q. Gao, “Wearable preimpact fall detector using SVM,” *Proc. Int. Conf. Sens. Technol. ICST*, pp. 1–6, 2016, doi: 10.1109/ICSensT.2016.7796223.
- [23] Y. Chen and Y. Xue, “A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer,” *Proc. - 2015 IEEE Int. Conf. Syst. Man, Cybern. SMC 2015*, pp. 1488–1492, 2016, doi: 10.1109/SMC.2015.263.
- [24] T. Zebin, P. J. Scully, and K. B. Ozanyan, “Human activity recognition with inertial sensors using a deep learning approach,” *Proc. IEEE Sensors*, no. October, 2017, doi: 10.1109/ICSENS.2016.7808590.
- [25] Y. Zhao and S. Zhou, “Wearable device-based gait recognition using angle embedded gait dynamic images and a convolutional neural network,” *Sensors (Switzerland)*, vol. 17, no. 3, 2017, doi: 10.3390/s17030478.
- [26] A. Murad and J. Y. Pyun, “Deep recurrent neural networks for human activity recognition,” *Sensors (Switzerland)*, vol. 17, no. 11, 2017, doi: 10.3390/s17112556.
- [27] Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, “LSTM Networks for Mobile Human Activity Recognition,” no. Icaita, pp. 50–53, 2016, doi: 10.2991/icaita-16.2016.13.
- [28] F. J. Ordóñez and D. Roggen, “Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition,” *Sensors (Switzerland)*, vol. 16, no. 1, 2016, doi: 10.3390/s16010115.
- [29] B. Almaslukh, B. Almaslukh, J. Almuhtadi, and A. Artoli, “An Effective Deep Autoencoder Approach for Online Smartphone-Based Human Activity Recognition,” *Int. J. Comput. Sci. Netw. Secur.*, vol. 17, no. 4, pp. 160–165, 2017, [Online]. Available: <https://www.researchgate.net/publication/323019783>.
- [30] A. Bosin, N. Dessì, and B. Pes, “Intelligent Data Engineering and

- Automated Learning - IDEAL 2007,” *Ideal*, vol. 4881, no. December, pp. 790–799, 2007, doi: 10.1007/978-3-540-77226-2.
- [31] Y. Zhang, Y. Zhang, Z. Zhang, J. Bao, and Y. Song, “Human activity recognition based on time series analysis using U-Net,” 2018, [Online]. Available: <http://arxiv.org/abs/1809.08113>.
- [32] C. C. Chang and C. J. Lin, “LIBSVM: A Library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–39, 2011, doi: 10.1145/1961189.1961199.
- [33] R. Couronné, P. Probst, and A. L. Boulesteix, “Random forest versus logistic regression: A large-scale benchmark experiment,” *BMC Bioinformatics*, vol. 19, no. 1, pp. 1–14, 2018, doi: 10.1186/s12859-018-2264-5.
- [34] F. Hernández, L. F. Suárez, J. Villamizar, and M. Altuve, “Human Activity Recognition on Smartphones Using a Bidirectional LSTM Network,” *2019 22nd Symp. Image, Signal Process. Artif. Vision, STSIVA 2019 - Conf. Proc.*, 2019, doi: 10.1109/STSIVA.2019.8730249.
- [35] O. Banos, J. M. Galvez, M. Damas, H. Pomares, and I. Rojas, “Window size impact in human activity recognition,” *Sensors (Switzerland)*, vol. 14, no. 4, pp. 6474–6499, 2014, doi: 10.3390/s140406474.
- [36] N. Zahradka, K. Verma, A. Behboodi, B. Bodt, H. Wright, and S. C. K. Lee, “An evaluation of three kinematic methods for gait event detection compared to the kinetic-based ‘gold standard,’” *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–15, 2020, doi: 10.3390/s20185272.
- [37] M. Czech and S. Patel, “GaitPy: An Open-Source Python Package for Gait Analysis Using an Accelerometer on the Lower Back,” *J. Open Source Softw.*, vol. 4, no. 43, p. 1778, 2019, doi: 10.21105/joss.01778.
- [38] S. Del Din, A. Godfrey, B. Galna, S. Lord, and L. Rochester, “Free-living gait characteristics in ageing and Parkinson’s disease: Impact of environment and ambulatory bout length,” *J. Neuroeng. Rehabil.*, vol. 13, no. 1, pp. 1–12, 2016, doi: 10.1186/s12984-016-0154-5.
- [39] H. T. T. Vu *et al.*, “A review of gait phase detection algorithms for lower limb prostheses,” *Sensors (Switzerland)*, vol. 20, no. 14, pp. 1–19, 2020, doi: 10.3390/s20143972.
- [40] A. June and R. Moe-nilssen, “A new method for evaluating motor control in gait under real-life environmental conditions . Part 1 : The instrument A new method for evaluating motor control in gait under

- real-life environmental conditions . Part 1 : The instrument,” vol. 13, no. 0268, pp. 320–327, 2014.
- [41] S. Del Din, A. Godfrey, and L. Rochester, “Validation of an Accelerometer to Quantify a Comprehensive Battery of Gait Characteristics in Healthy Older Adults and Parkinson’s Disease: Toward Clinical and at Home Use,” *IEEE J. Biomed. Heal. Informatics*, vol. 20, no. 3, pp. 838–847, 2016, doi: 10.1109/JBHI.2015.2419317.
- [42] J. McCamley, M. Donati, E. Grimpampi, and C. Mazzà, “An enhanced estimate of initial contact and final contact instants of time using lower trunk inertial sensor data,” *Gait Posture*, vol. 36, no. 2, pp. 316–318, 2012, doi: 10.1016/j.gaitpost.2012.02.019.
- [43] W. Zijlstra and A. L. Hof, “Assessment of spatio-temporal gait parameters from trunk accelerations during human walking,” *Gait Posture*, vol. 18, no. 2, pp. 1–10, 2003, doi: 10.1016/S0966-6362(02)00190-X.
- [44] S. Fritz and M. Lusardi, “White paper: ‘walking speed: The sixth vital sign,’” *J. Geriatr. Phys. Ther.*, vol. 32, no. 2, pp. 2–5, 2009, doi: 10.1519/00139143-200932020-00002.
- [45] A. Tura, M. Raggi, L. Rocchi, A. G. Cutti, and L. Chiari, “Gait symmetry and regularity in transfemoral amputees assessed by trunk accelerations,” *J. Neuroeng. Rehabil.*, vol. 7, no. 1, pp. 1–10, 2010, doi: 10.1186/1743-0003-7-4.
- [46] R. Moe-Nilssen and J. L. Helbostad, “Estimation of gait cycle characteristics by trunk accelerometry,” *J. Biomech.*, vol. 37, no. 1, pp. 121–126, 2004, doi: 10.1016/S0021-9290(03)00233-1.
- [47] B. Fasel *et al.*, “A wrist sensor and algorithm to determine instantaneous walking cadence and speed in daily life walking,” *Med. Biol. Eng. Comput.*, vol. 55, no. 10, pp. 1773–1785, 2017, doi: 10.1007/s11517-017-1621-2.
- [48] M. Vondra and R. Vích, “Adaptive Comb Filtering in Speech Enhancement,” no. 3, pp. 3–5.
- [49] T. D. A. Busch *et al.*, “Factors associated with lower gait speed among the elderly living in a developing country: A cross-sectional population-based study,” *BMC Geriatr.*, vol. 15, no. 1, pp. 1–9, 2015, doi: 10.1186/s12877-015-0031-2.
- [50] J. C. Brown, M. O. Harhay, and M. N. Harhay, “Walking cadence and

mortality among community-dwelling older adults,” *J. Gen. Intern. Med.*, vol. 29, no. 9, pp. 1263–1269, 2014, doi: 10.1007/s11606-014-2926-6.

- [51] A. Zijlstra, E. D. de Bruin, N. Bruins, and W. Zijlstra, “The step length - Frequency relationship in physically active community-dwelling older women,” *Eur. J. Appl. Physiol.*, vol. 104, no. 3, pp. 427–434, 2008, doi: 10.1007/s00421-008-0795-6.
- [52] T. Kimura, H. Kobayashi, E. Nakayama, and M. Hanaoka, “Effects of aging on gait patterns in the healthy elderly,” *Anthropol. Sci.*, vol. 115, no. 1, pp. 67–72, 2007, doi: 10.1537/ase.060309.
- [53] M. Y. Osoba, A. K. Rao, S. K. Agrawal, and A. K. Lalwani, “Balance and gait in the elderly: A contemporary review,” *Laryngoscope Investig. Otolaryngol.*, vol. 4, no. 1, pp. 143–153, 2019, doi: 10.1002/lio2.252.
- [54] G. Yogev, M. Plotnik, C. Peretz, N. Giladi, and J. M. Hausdorff, “Gait asymmetry in patients with Parkinson’s disease and elderly fallers: When does the bilateral coordination of gait require attention?,” *Exp. Brain Res.*, vol. 177, no. 3, pp. 336–346, 2007, doi: 10.1007/s00221-006-0676-3.