



**Politecnico  
di Torino**

# **POLITECNICO DI TORINO**

**Laurea Magistrale in Ingegneria Informatica**

**Aprile 2023**

## **Monitoraggio della motilità notturna su soggetti con Parkinson tramite sensori inerziali: un approccio di Supervised Learning**

**Relatrici**

**Prof.ssa Gabriella Olmo**

**Ing.ra Irene Rechichi**

**Candidato**

**Luca Di Gangi**



*"It'll pass"*





# Sommario

La Malattia di Parkinson è un disturbo neurodegenerativo che coinvolge funzioni quali il controllo del movimento e dell'equilibrio e ad oggi risulta essere la malattia più frequente tra i disturbi del movimento. Il Parkinson è diffuso in tutto il mondo e coinvolge tutti i gruppi etnici, mostrando riscontro in entrambi i sessi, con una lieve prevalenza verso i soggetti di sesso maschile. Secondo le statistiche raccolte in questi ultimi anni, l'età media di esordio risulta essere intorno ai 58-60 anni dei quali circa il 5% dei pazienti può presentare un esordio giovanile già tra i 21 ed i 40 anni, mentre la comparsa prima del compimento dei 20 anni si è dimostrata essere particolarmente rara. L'incidenza registrata sulla popolazione over 60 si aggira sull' 1-2% , sino a raggiungere il 3-5% nei casi in cui l'età sia superiore agli 85 anni. Tra tutti i disturbi che comporta la Malattia di Parkinson, i disturbi del sonno sono la complicità che influisce maggiormente sulla salute fisica e mentale dei soggetti affetti da Parkinson, nello specifico le Ipocinesie e le Acinesie. Partendo da un campione rappresentativo costituito da circa trenta soggetti, suddivisi in due gruppi, di cui uno di controllo e uno formato da pazienti con diagnosi di Parkinson, di ognuno è stata monitorata l'attività motoria durante il sonno attraverso l'applicazione di un dispositivo che raccogliesse i parametri generati da accelerometro, giroscopio e magnetometro. Successivamente, partendo dai dati raccolti con sensori inerziali, sono state estrapolate delle features con l'obiettivo di fornire un'analisi quantitativa della motilità notturna. Al fine di effettuare un'analisi completa dal punto di vista sia quantitativo sia qualitativo, si è tenuto conto anche della percezione del sonno da parte dei soggetti in analisi. A tal proposito, sono stati somministrati due questionari: short-Pittsburg Sleep Quality Index (sPSQI) e Sleep Survey (SLEEPS), il primo comunemente utilizzato nella pratica clinica, dai quali si è ottenuto un punteggio che è stato utilizzato come feature.

Su tutte le features raccolte è stato effettuato un processo di selezione, al fine di mantenere soltanto quelle con una buona correlazione rispetto alla diagnosi di Parkinson. Sulle features rimanenti sono stati implementati quattro classificatori, Support Vector Machine, K-Nearest Neighbour, Random Forest e Gradient Boosting, sui quali è stata effettuata un'ottimizzazione degli iperparametri e successivamente

un confronto per identificare il modello ottimale.

Infine, è stata proposta una soluzione di medicina pervasiva per l'home care con dispositivi wearable al fine di rendere possibile un continuo monitoraggio permettendo una maggiore accessibilità alle cure. Grazie a questo studio si sono riscontrate buone correlazioni tra diversi aspetti della motilità notturna e la diagnosi di Parkinson, inoltre, dalle analisi statistiche effettuate sui dati raccolti, si sono riscontrati ottimi risultati in termini di omogeneità e di normalità, dando così conferma della validità del lavoro eseguito.

# Elenco delle tabelle

1.1	Sintomi malattia di Parkinson . . . . .	3
1.2	UK PD Brain Bank Criteria for the diagnosis of Parkinson's disease	4
1.3	Parkinson's Disease Sleep Scale adattata da Chaudhuri [19] . . . . .	6
3.1	Domande sPSQI . . . . .	27
3.2	Calcolo score sPSQI . . . . .	28
3.3	Domande SLEEPS . . . . .	29
4.1	Coefficienti di Correlazione di Pearson tra features e diagnosi di Parkinson . . . . .	40
4.2	Dizionario per ottimizzazione Iperparametri . . . . .	42
4.3	Risultato dell'ottimizzazione degli Iperparametri . . . . .	42
4.4	Risultati Classificazione . . . . .	42
4.5	Statistiche descrittive Sani-PD . . . . .	45
4.6	Test di Normalità di Shapiro-Wilk su Sani-PD . . . . .	46
4.7	Test T di Student su Sani-PD . . . . .	47
4.8	Test U di Mann-Whitney su Sani-PD . . . . .	48
4.9	Coefficienti di Correlazione di Pearson tra features e diagnosi di Parkinson . . . . .	51
4.10	Risultato dell'ottimizzazione degli Iperparametri . . . . .	51
4.11	Risultati Classificazione Good Sleepers . . . . .	52
4.12	Statistiche descrittive Good/Bad Sleepers . . . . .	54
4.13	Test di Normalità di Shapiro-Wilk su Good/Bad Sleepers . . . . .	55
4.14	Test T di Student su Good/Bad Sleepers . . . . .	56
4.15	Test U di Mann-Whitney su Good/Bad Sleepers . . . . .	57

# Elenco delle figure

1.1	Confronto tra substantia nigra sana e substantia nigra ridotta dovuta alla presenza del morbo di Parkinson . . . . .	2
2.1	Illustrazione di tre delle coordinate spaziali discusse precedentemente: n-frame in un punto localizzato sulla terra, e-frame ruota con la terra e i-frame . . . . .	8
2.2	Accelerometer system: a) electromechanical schematic; b) equivalent system model; and c) simplified sensor interface schematic . . . . .	9
2.3	Gyroscope system: a) electromechanical schematic; b) equivalent system model; c) simplified sensor interface schematic . . . . .	11
2.4	(a) Schema delle linee di forza del campo magnetico (verdi) della terra (blu). (b) Schema di una parte della terra in cui il campo magnetico locale $m^n$ crea un angolo $\delta$ con il piano orizzontale. . . .	12
2.5	Iperpiano che separa i support vectors corrispondenti ad ognuna delle due classi da predirre . . . . .	14
2.6	Histogram Approach . . . . .	16
2.7	Illustrazione dell'estimatore di densità K-NN . . . . .	17
2.8	Grafico con 200 data points che mostra i valori di $x_6$ raffigurati su $x_7$	18
2.9	Spazio di input a due dimensioni partizionato. adapted from [22] . .	18
2.10	Binary Tree. adapted from [22] . . . . .	19
2.11	25° e 75° Percentile in una distribuzione Normale . . . . .	23
2.12	Matrice di confusione . . . . .	24
3.1	Sensore Shimmer3 . . . . .	30
3.2	Shimmer3 core component diagram . . . . .	31
3.3	Shimmer3 mainboard layout (a sinistra vista dall'alto, a destra dal basso) . . . . .	31
3.4	Shimmer3 indossato . . . . .	32
3.5	BoxPlot con distribuzione dell'età nei soggetti sani e soggetti con diagnosi di Parkinson . . . . .	33

3.6	Istogramma rappresentante il numero di soggetti e il corrispondente sesso, suddiviso tra soggetti sani e soggetti con diagnosi di Parkinson	33
3.7	Funzionamento filtro passabasso . . . . .	34
3.8	Accelerometro . . . . .	35
3.9	Giroscopio . . . . .	35
3.10	Magnetometro . . . . .	36
4.1	Diagramma Entità-Relazione per Database . . . . .	58
4.2	Prototipo di funzionamento app per Apple Watch . . . . .	59
4.3	Mockup Home . . . . .	60
4.4	Mockup Schermata paziente . . . . .	60
4.5	Mockup schermata dettaglio . . . . .	61
5.1	Risultati statistiche delle features maggiormente rilevanti rispetto la classe Sani/PD . . . . .	63
5.2	Risultati statistiche delle features maggiormente rilevanti rispetto la classe Bad/Good Sleepers . . . . .	64



# Acronyms

**PD**

Parkinson's Disease, Malattia di Parkinson

**SVC**

Support Vector Classifier

**K-NN**

K-Nearest Neighbour

**RF**

Random Forest

**GB**

Gradient Boosting

**sPSQI**

short Pittsburg Sleep Quality Index

**SLEEPS**

Sleep Survey

# Capitolo 1

## Introduzione

### 1.1 Malattia di Parkinson

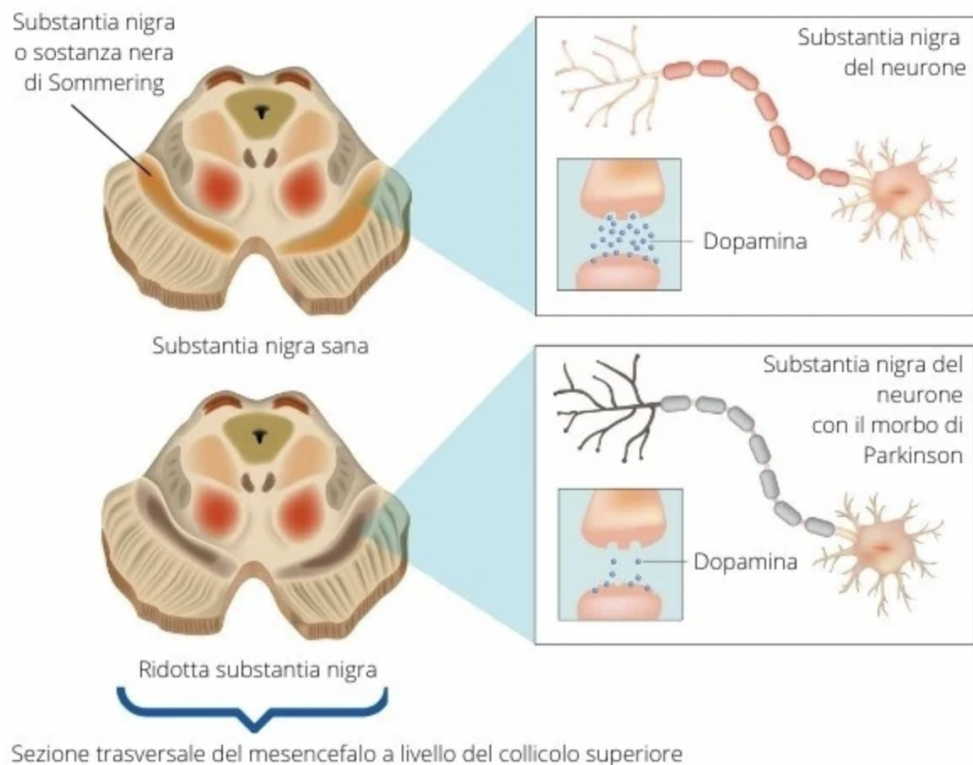
La Malattia di Parkinson è la seconda malattia neurodegenerativa più comune dopo l'Alzheimer [1]. questa sindrome venne descritta da James Parkinson nel 1817 come "Movimento involontario tremolante, con diminuzione della potenza muscolare, nelle parti non in azione e anche quando sono sostenute; con una propensione a piegare il tronco in avanti, e a cambiare da un passo di camminata a uno di corsa" [2]. Ad oggi il morbo di Parkinson viene definito come malattia idiopatica del sistema nervoso, ossia una patologia non dovuta a cause esterne note [3], caratterizzata da manifestazioni sia del sistema motorio sia del sistema non motorio [4]. È un disturbo cronico e neurodegenerativo che si presenta maggiormente in persone di età avanzata, ma può verificarsi anche in soggetti più giovani. Secondo le statistiche effettuate nel 2013[5], gli uomini rispetto alle donne hanno una maggiore tendenza a sviluppare la Malattia di Parkinson di 1.5 volte. Per quanto riguarda l'età, anche questa risulta essere un fattore determinante, infatti sempre dalle stesse statistiche risulta che nel ventennio successivo al compimento dei sessantacinque anni si ha un drastico incremento della probabilità di sviluppare questo disturbo.

#### 1.1.1 Fisiopatologia del Parkinson

Per comprendere al meglio la patologia è bene affrontarne l'aspetto fisiopatologico al fine di poter avere un'idea complessiva di quello che avviene a livello biologico prima ancora di affrontare la sintomatologia. La Malattia di Parkinson è un disturbo del sistema extrapiramidale, ossia una patologia che interessa i nuclei della base, un gruppo di cellule nervose localizzate in profondità nel cervello facenti parte del sistema extrapiramidale [6], questa malattia è caratterizzata dalla perdita della funzione dopaminergica che può comportare la perdita di funzioni motorie [7], questo si verifica poiché la progressiva degenerazione dei neuroni



dopaminergici nella substantia nigra pars compacta (SNpc) causa appunto la perdita della funzione dopaminergica, nello specifico si verifica una complessa interazione di aggregazione di  $\alpha$ -sinucleina, disfunzione dei mitocondri, lisosomi o del trasporto di vescicole, problemi di trasporto sinaptico e neuroinfiammazione. [8] La perdita di cellule dopaminergiche nigrostriatali provoca un gradiente di dopamina striatale, producendo uno squilibrio tra le vie dirette (facilitanti) e indirette (inibitorie) attraverso i gangli basali, con conseguente bradicinesia.[9]



**Figura 1.1:** Confronto tra substantia nigra sana e substantia nigra ridotta dovuta alla presenza del morbo di Parkinson

### 1.1.2 Sintomi del Parkinson

Si stima che più dell'80% delle cellule dopaminergiche del sistema nigro-striatale vengano perse prima dell'effettiva comparsa dei sintomi motori.[10] Generalmente, infatti, questa malattia viene diagnosticata soltanto dopo la comparsa dei primi sintomi. Nell'articolo pubblicato da J Jankovic [11], per indicare i quattro principali sintomi di questo disturbo, viene utilizzato l'acronimo TRAP, che rispettivamente indica *Tremor at rest*, *Rigidity*, *Akinesia (or bradykinesia)* and *Postural instability*.

A causa della diversità negli stili di vita di ogni persona affetta da PD, i sintomi motori e non motori devono essere valutati in modo differente, motivo per cui sono state introdotte diverse scale al fine di determinare il livello di riduzione di motilità e del livello di disabilità raggiunto dai pazienti Parkinson, tuttavia non tutte le scale rese disponibili sono state validate e di conseguenza non possono essere ritenute tutte affidabili. La Unified Parkinson's Disease Rating scale (UPDRS) ad oggi risulta essere la più famosa e attendibile, grazie a questa scala infatti si è potuto osservare come il livello di deterioramento non è lineare, si è invece notato come nella fase iniziale della malattia si abbia un deterioramento più veloce. Sempre grazie alla UPDRS è stato dimostrato come nei pazienti di età più giovane ci sia un maggiore rischio nel riscontrare bradicinesie in seguito all'assunzione di Levodopa (il farmaco maggiormente utilizzato per il rallentamento del deterioramento neurale). [12]

Sintomi motori	Sintomi non motori
Tremor, bradykinesia, rigidity, postural instability	Cognitive impairment, bradyphrenia, tip-of-the-tongue (word finding) phenomenon
Hypomimia, dysarthria, dysphagia, sialorrhoea	b+Depression, apathy, anhedonia, fatigue, other behavioural and psychiatric problems
Decreased arm swing, shuffling gait, festination difficulty arising from chair, turning in bed	Sensory symptoms: anosmia, ageusia, pain (shoulder, back), paresthesias
Micrographia, cutting food, feeding, hygiene, slow activities of daily living	Dysautonomia (orthostatic hypotension, constipation, urinary and sexual dysfunction, abnormal sweating, seborrhoea), weight loss
Glabellar reflex, blepharospasm, dystonia, striatal deformity, scoliosis, camptocormia	Sleep disorders (REM behaviour disorder, vivid dreams, daytime drowsiness, sleep fragmentation, restless legs syndrome)

**Tabella 1.1:** Sintomi malattia di Parkinson

Stando a quanto affermato prima, poichè la diagnosi è basata sulla sintomatologia, nella fase diagnostica vengono seguiti dei criteri definiti dalla UK PD Brain Bank [13] mostrata nella tabella 1.2. Uno dei sintomi principali è sicuramente la Bradicinesia, ossia il rallentamento di movimenti volontari con una progressiva riduzione in termini di velocità e ampiezza di azioni ripetitive, che si può tradurre in rigidità muscolare. Altri sintomi comuni sono tremori in situazioni di riposo e instabilità posturale. La bradicinesia inoltre può comportare una diminuzione di espressività facciale, anche

detta ipomimia, e una riduzione dell'ampiezza della propria calligrafia, anche detta micrografia. L'80% dei pazienti presenta tremore agli arti

---

**Step 1:** Diagnosis of Parkinsonian syndrome Bradykinesia (slowness of initiation of voluntary movement with progressive reduction in speed and amplitude of repetitive action) And at least one of the following

- Muscular rigidity
- 4-6 Hz rest tremor
- Postural instability not caused by primary visual, vestibular, cerebellar or proprioceptive dysfunction

**Step 2:** Exclusion criteria for Parkinson's disease

- History of repeated strokes with stepwise progression of parkinsonian features
- History of repeated head injury
- History of definite encephalitis
- Oculogyric crises
- Neuroleptic treatment at onset of symptoms
- More than one affected relative
- Sustained remission
- Strictly unilateral features after 3 years
- Supra-nuclear gaze palsy
- Cerebellar signs
- Early severe autonomic involvement
- Early severe dementia with disturbances of memory, language and praxis
- Babinski sign
- Presence of cerebral tumour or communicating hydrocephalus on CT scan
- Negative response to large doses of levodopa (if malabsorption excluded)
- MPTP exposure

**Step 3:** supportive prospective positive criteria for Parkinson's disease (three or more required for diagnosis of definite Parkinson's disease)

- Unilateral onset
  - Rest tremor present
  - Progressive disorder
  - Persistent asymmetry affecting side of onset more
  - Excellent response (70-100% ) to levodopa
  - Severe levodopa-induced chorea
  - Levodopa response for 5 years or more
  - Clinical course of 10 years or more
-

**Tabella 1.2:** UK PD Brain Bank Criteria for the diagnosis of Parkinson’s disease

### 1.1.3 Disturbi del sonno: Ipocinesie e Acinesie

Un questionario somministrato nel 1988 [14] mette in risalto come il 98% dei pazienti affetti da PD abbiano riscontrato disturbi notturni o durante il risveglio, inoltre fino al 30% dei pazienti presentava problemi nella regolazione della veglia. Dietro al danneggiamento delle funzioni dei neurotrasmettitori sono presenti altre cause che contribuiscono nell’alterazione del sonno, tra queste abbiamo l’utilizzo di farmaci dopaminergici, noti per influenzare le regolazioni sonno-veglia. Anche l’assunzione di altri farmaci spesso utilizzati in soggetti più anziani e nella popolazione multimorbosa può implicare sintomi PD che incidano sul sonno dei pazienti, con disturbi quali acinesie e ipocinesie. La presenza di acinesie notturne, note per essere manifestazioni motorie del Parkinson, diminuisce significativamente la capacità di effettuare movimenti in maniera confortevole e può portare all’interruzione del sonno. Inoltre alcuni pazienti riportano la necessità di muoversi dovuta ad acinesie/ipocinesie.[15] Negli studi di JP Lakke [16] viene mostrato come la rotazione assiale sia maggiormente alterata nei casi in cui il soggetto è sdraiato rispetto a quando è in piedi. Questo può incrementare sintomi causati dalle ipocinesie a riposo o durante la notte, simili ai sintomi della sindrome delle gambe senza riposo, un disturbo neurologico in cui il paziente percepisce la necessità di muovere gli arti inferiori per trovare conforto a seguito di fastidio e fitte alle gambe. A partire dal 2002 è stata resa disponibile la scala verificata Parkinson’s Disease Sleep Scale (PDSS) [17], disponibile in cinque lingue, il primo strumento formale per quantificare i disturbi del sonno legati alla malattia di Parkinson. La scala PDSS si è dimostrata robusta e con un buon potere discriminatorio tra i pazienti di controllo e i pazienti sperimentali. Più recentemente è stata resa disponibile la scala SCOPA-SLEEP, una scala suddivisa in due parti che descrive il sonno notturno e il livello di sonnolenza durante il giorno. [18]

---

**Parkinson's Disease Sleep Scale (PDSS)**

---

How would you rate the following, based on your experience during the past one week.

1. The overall quality of your night's sleep is:
  2. Do you have difficulty falling asleep each night?
  3. Do you have difficulty staying asleep?
  4. Do you have restlessness of legs or arms at night or in the evening causing disruption of sleep?
  5. Do you fidget in bed?
  6. Do you suffer from distressing dreams at night?
  7. Do you suffer from distressing hallucinations at night (seeing or hearing things that you are told do not exist)?
  8. Do you get up at night to pass urine?
  9. Do you have incontinence of urine because you are unable to move due to "off" symptoms?
  10. Do you experience numbness or tingling of your arms or legs which wake you from sleep at night?
  11. Do you have painful muscle cramps in your arms or legs whilst sleeping at night?
  12. Do you wake early in the morning with painful posturing of arms or legs?
  13. On waking do you experience tremor?
  14. Do you feel tired and sleepy after waking in the morning?
  15. Have you unexpectedly fallen asleep during the day?
- 

**Tabella 1.3:** Parkinson's Disease Sleep Scale adattata da Chaudhuri [19]

# Capitolo 2

## Stato dell'arte

### 2.1 Sensori inerziali

Negli ultimi anni, a seguito della rivoluzione tecnologica, i sensori inerziali hanno iniziato ad essere presenti in sempre più dispositivi, sino ad essere integrati in oggetti facenti parte della quotidianità della maggior parte degli utenti. Basti pensare agli smartphone, ai tablet, ai sistemi di gioco, e non solo. A portare questa grande diffusione vi è la rapida diminuzione negli anni del costo dei sistemi micro elettromeccanici (MEMS) e semiconduttori complementari ad ossidi metallici (CMOS). Grazie all'integrazione dei MEMS/CMOS, in un singolo substrato di silicone possono coesistere tanti tipi di sensori multiasse, tutto racchiuso in piccoli contenitori a basso costo. In questo elaborato verranno trattati Accelerometro, Giroscopio e Magnetometro, in quanto integrati nel sensore utilizzato nella ricerca (che verrà a sua volta trattato nel successivo capitolo). [20]

#### 2.1.1 Coordinate spaziali

Al fine di discurre nel dettaglio le quantità misurate dai sopracitati sensori inerziali, è bene introdurre le seguenti coordinate spaziali:

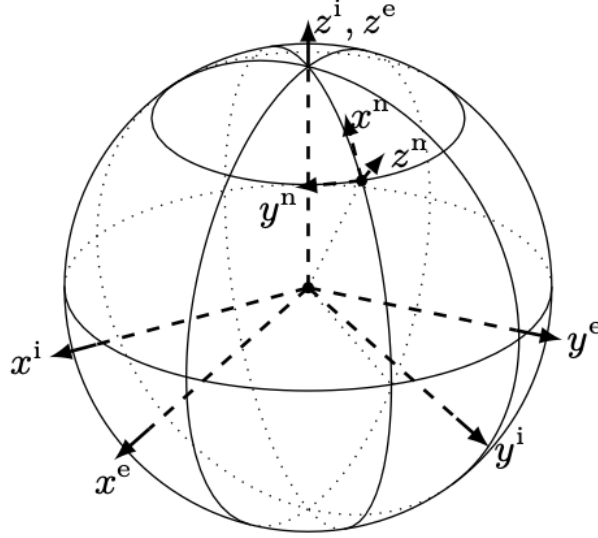
- **Body frame  $b$ :** sono le coordinate del frame del movimento IMU. L'origine è localizzato al centro della triade dell'accelerometro ed è allineato all'involucro. Tutte le misurazioni inerziali fanno riferimento a questo frame.
- **Navigation frame  $n$ :** è un frame geografico locale in cui vogliamo navigare, ossia ciò a cui si fa riferimento è la posizione del frame  $b$  rispetto questo frame. Nella maggior parte degli utilizzi risulta essere stazionario rispetto alla terra, tuttavia nei casi in cui per il sensore è prevista il tracciamento di attività su lunghe distanze, il frame  $n$  può muoversi e ruotare.

- **Inertial frame  $i$ :** è stazionario, l'IMU misura l'accelerazione lineare e la velocità angolare rispetto a questo frame. L'origine di riferimento è posizionato al centro della Terra.
- **Earth frame  $e$ :** coincide con il frame  $i$  ma ruota con la Terra. Il suo origine è posizionato al centro della terra assieme ai suoi assi, allineati con essa.

$b$  verrà espresso come  $R = (R^{nb})^T$ . Di conseguenza il vettore  $x$  nel body frame viene espresso come  $(x^b)$  e nel navigation frame come  $(x^n)$ , da cui otteniamo le seguenti relazioni:

$$x^n = R^{nb} x^b$$

$$x^b = (R^{nb})^T x^n = R^{bn} x^n$$



**Figura 2.1:** Illustrazione di tre delle coordinate spaziali discusse precedentemente: n-frame in un punto localizzato sulla terra, e-frame ruota con la terra e i-frame

### 2.1.2 Accelerometro

L'accelerometro viene utilizzato per determinare l'accelerazione del sistema in cui si trova, spesso misurata in  $g$ , in cui  $1 g$  equivale alla forza di gravità del pianeta Terra, ossia  $9.81 m/s^2$ . Scendendo nei dettagli, l'accelerometro misura la forza specifica  $f$  di un corpo  $b$ , la quale può essere espressa come

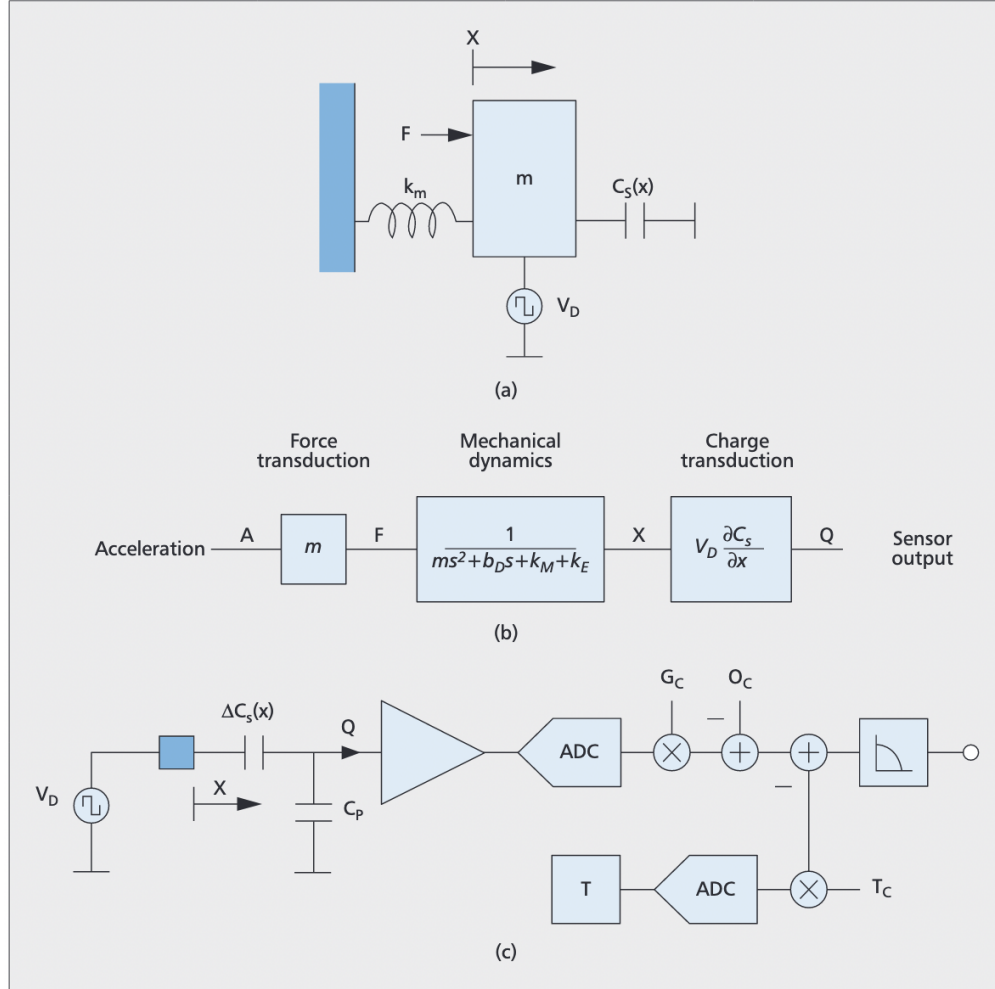
$$f^b = R^{bn}(a_{ii}^n - g^n)$$

in cui come espresso prima  $g$  indica il vettore della gravità terrestre, mentre  $a_{ii}^n$  denota l'accelerazione lineare del sensore, che può essere espressa come

$$a_{ii}^n = R^{ne} R^{ei} a_{ii}^i$$

$f_t^b$  ad ogni istante di tempo  $t$ , ricavandone così .

Nella figura 2.2 (a) viene mostrata una versione semplificata dello schema elet-



**Figura 2.2:** Accelerometer system: a) electromechanical schematic; b) equivalent system model; and c) simplified sensor interface schematic

tromeccanico di un accelerometro ed una versione semplificata del circuito dello stesso. L'accelerazione viene generata dal movimento del sensore o dalla gravità. Sempre nella stessa figura, nella sezione (b) viene mostrato lo schema equivalente precedentemente descritto, sottolineando come la trasduzione dell'accelerazione in



forza attiva sulla massa avvenga in virtù della massa stessa. Rimanendo sulla stessa figura, abbiamo infine nella sezione (c) uno schema semplificato dell'interfaccia del sensore in questione. Il flusso di carica dell'amplificatore è direttamente proporzionale all'accelerazione con un offset minimo. Poichè la capacità di rilevamento è quasi statica, per misurare l'accelerazione occorre l'utilizzo di uno stimolo AC. Per un approccio a tempo discreto, viene applicata una tensione di azionamento,  $V_D$ , che applicata alla massa di prova genera una carica che viene accumulata dall'amplificatore e successivamente digitalizzata dal convertitore analogico/digitale (ADC)

### 2.1.3 Giroscopio

Attraverso il giroscopio viene misurata la velocità angolare di rotazione, per la quale l'unità di misura più utilizzata sono i gradi per secondo (*dps*). Il funzionamento del giroscopio si basa sul rilevamento dell'accelerazione di Coriolis su una massa vibrante in proporzione alla velocità di rotazione lungo un'asse ortogonale all'asse di vibrazione. Misurando l'accelerazione di Coriolis si può calcolare la velocità di angolare in quanto l'ampiezza della velocità della massa di prova è nota. La velocità angolare viene indicata come  $\omega_{ib}^b$  e può essere espressa come:

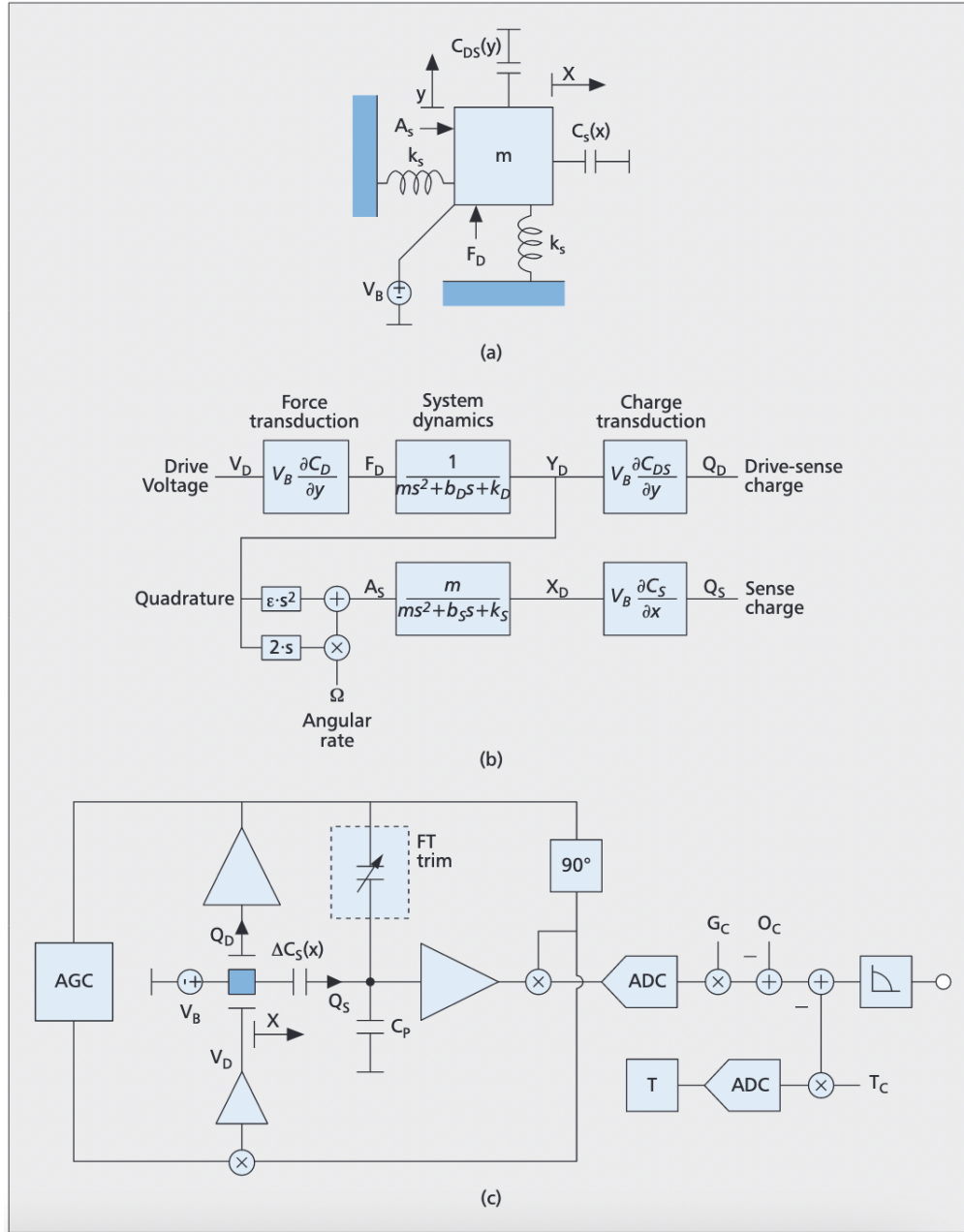
$$\omega_{ib}^b = R^{bn}(\omega_{ie}^n + \omega_{en}^n) + \omega_{nb}^b$$

in cui  $R^{bn}$  è la matrice di rotazione dal navigation frame al body frame.  $\omega_{ie}$  indica la velocità angolare dell'*earth frame*, ossia dell Terra, parametro noto che corrisponde approssimativamente a  $7.29 \cdot 10^{-5} rad/s$ .

Nella figura 2.3 è illustrata una versione base di un sistema a giroscopio, incluso lo schema elettromeccanico, un sistema a modello equivalente e un'interfaccia del diagramma semplificato. Come mostrato nella sezione (a), il sistema meccanico possiede due gradi di libertà, inclusi un asse di rilevamento (x) ed un asse di azionamento (y). Il moto della massa di prova viene rappresentato attraverso una sinusoide.

### 2.1.4 Magnetometro

Il magnetometro permette di misurare il campo magnetico locale, che consiste nell'insieme del campo magnetico della Terra e il campo generato da un oggetto con proprietà magnetiche. Il campo magnetico locale  $m^n$  è illustrato nella figura 2.4, la sua componente orizzontale punta verso il polo nord magnetico della terra. Il rapporto tra la componente orizzontale e quella verticale può essere chiamato *dip angle* ( $\delta$ ). Assumendo che il sensore non copra distanze significative in relazione alla dimensione della Terra, il campo magnetico locale può essere semplificato come costante. Il magnetometro inoltre può dare informazioni su qualsiasi posizione



**Figura 2.3:** Gyroscope system: a) electromechanical schematic; b) equivalent system model; c) simplified sensor interface schematic

sulla Terra ad eccezione dei poli magnetici, sui quali il campo magnetico locale  $m^n$  risulta essere locale, come appunto deducibile dalle linee di forza rappresentate in figura 2.4(a). La grandezza del campo magnetico è trascurabile, motivo per cui il

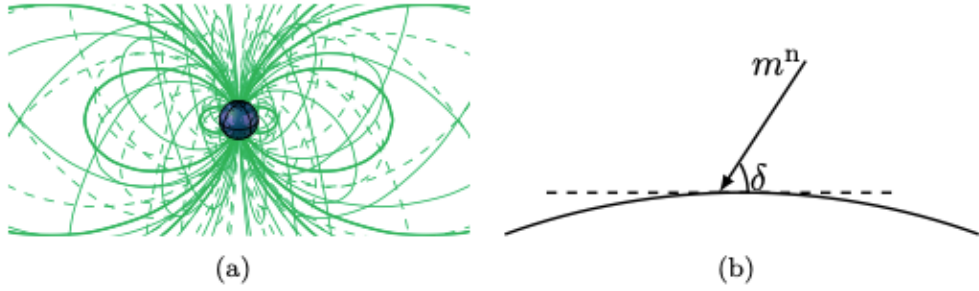
campo magnetico della Terra può essere descritto come

$$m^n = (\cos\delta \ 0 \ \sin\delta)^T$$

di conseguenza può essere affermato che  $\|m^n\|_2 = 1$ . Considerando che il magnetometro misura solamente il campo magnetico locale, le sue misurazioni  $y_{m,t}$  possono essere rappresentate come

$$y_{m,t} = R_t^{bn} m^n + e_{m,t}$$

in cui il rumore  $e_{m,t}$  rappresenta la misura del rumore percepita dal magnetometro, data appunto dall'incertezza.



**Figura 2.4:** (a) Schema delle linee di forza del campo magnetico (verdi) della terra (blu). (b) Schema di una parte della terra in cui il campo magnetico locale  $m^n$  crea un angolo  $\delta$  con il piano orizzontale.

[21]

## 2.2 Modelli di apprendimento supervisionato

La classificazione [22] è un forma di machine learning in cui vengono mappati i dati di input in dati di output basandosi su un esempio di input-output sviluppato in una prima fase di allenamento. Attraverso la classificazione, alcune features correlate ad un set di osservazioni d'esempio possono essere usate per allenare una funzione decisionale che suddivida in classi fornendo una relativa accuratezza. In questa sezione verranno approfonditi i principali modelli di apprendimento supervisionato utilizzati nel corso di questo studio, ossia, in ordine di presentazione: Support Vector Machines, K-Nearest Neighbour, Random Forest e Gradient Boosting.

### 2.2.1 Margine Massimo

Spesso le Support Vector Machines vengono generalizzate con i classificatori a margine massimo. Il concetto alla base di questo classificatore è la presenza di

un iperpiano che permette di classificare un insieme di dati separandolo in due o più gruppi. Per meglio comprendere il funzionamento di questo modello è bene definire cosa sia un **Iperpiano**. In uno spazio ad  $n$ -dimensioni, un **Iperpiano** è un sottospazio affine piatto con dimensione  $n-1$ . Questo implica che in uno spazio a 2-dimensioni, l'iperpiano sarà una linea dritta che separa lo spazio in due metà. L'iperspazio è definito dalla seguente equazione:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$$

i punti che si trovano sull'iperpiano devono rispettare la suscritta equazione. Tuttavia possono esistere anche regioni al di sopra e al di sotto :

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 > 0$$

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 < 0$$

La rappresentazione matematica del classificatore a margine massimo può essere rappresentata dal seguente problema di ottimizzazione:

$$\text{funzione obiettivo : } \underset{\beta_0, \beta_1, \dots, \beta_n}{\text{maximize}} M$$

$$\text{vincolo 1 : } \text{soggetto a } \sum_{j=1}^n \beta_j^2 = 1$$

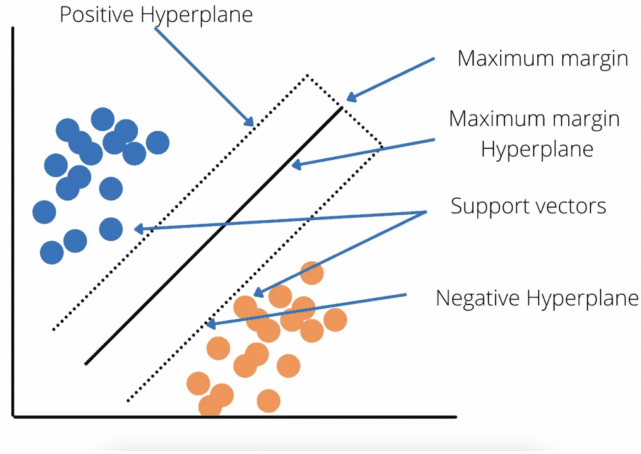
$$\text{vincolo 2 : } y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in}) \geq M \forall i = 1, \dots, m$$

$$\text{class labels : } y_1, \dots, y_m \in \{-1, 1\}$$

il secondo vincolo assicura che l'osservazione sarà sul lato corretto dell'iperpiano prendendo il prodotto dei coefficienti con le variabili  $x$  e inf fine con in indicatore a variabile di classe. Come mostrato in figura 2.5, si possono tracciare infiniti iperpiani per separare le due classi (rappresentate dai due colori diversi). Tuttavia, il classificatore a margine massimo ha come obiettivo quello di massimizzare il margine tra l'iperpiano positivo e quello negativo, sino a toccare gli estremi, chiamati *support vectors*. Le performance dei classificatori dipendono dai support vectors.

## 2.2.2 Support vector classifier

I classificatori a Support Vector sono una versione estesa dei classificatori a margine massimo, in cui però sono tollerate alcune violazioni per i casi non separabili, al fine di poter creare un migliore adattamento. Nei casi reali difficilmente si hanno dati con classi che possono essere separate perfettamente, molte classi hanno alcune caratteristiche che si sovrappongono con quelle di altre classi. La rappresentazione



**Figura 2.5:** Iperpiano che separa i support vectors corrispondenti ad ognuna delle due classi da predire

matematica del classificatore a Support vector riprende quella del classificatore a margine massimo con alcune piccole correzioni sui vincoli:

*funzione obiettivo :*  $\underset{\beta_0, \beta_1, \dots, \beta_n}{\text{maximize}} M$

*vincolo 1 :*  $\text{soggetto a } \sum_{j=1}^n \beta_j^2 = 1$

*vincolo 2 :*  $y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in}) \geq M(1 - \epsilon_i) \forall i = 1, \dots, m$

*vincolo 3 :*  $\epsilon_i \geq 0, \sum_{i=1}^n \epsilon_i \leq C$

*class labels :*  $y_1, \dots, y_m \in \{-1, 1\}$

Nel vincolo 4 , il valore C è un parametro non negativo che permette di gestire gli errori nel modello. Con un valore alto di C si può avere un modello più robusto, mentre con un valore piccolo di C si ottiene un modello più flessibile. In sostanza il valore C permette di regolare il modello di machine learning in utilizzo.

### C-Support Vector Classification

Dati i training vectors  $\mathbf{x}_i \in R^n, i = 1, \dots, l$ , in due classi, e un indicator vector  $\mathbf{y} \in R^l$  tale per cui  $y_i \in 1, -1$ , C-SVC risolve il seguente problema di ottimizzazione:

$$\min_{\omega, b, \xi} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \xi_i$$

$$\begin{aligned} y_i(\boldsymbol{\omega}^T \phi(\mathbf{x}_i) + b) &\geq 1 - \xi_i \\ \xi_i &\geq 0, i = 1, \dots, l \end{aligned}$$

in cui  $\phi(\mathbf{x}_i)$  mappa  $\mathbf{x}_i$  in uno spazio ad una dimensione superiore e  $C > 0$  è il parametro di regolarizzazione. A causa della possibile alta dimensionalità del vettore  $\boldsymbol{\omega}$ , solitamente viene risolto il seguente problema:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \frac{1}{2} \boldsymbol{\alpha}^T Q \boldsymbol{\alpha} - \mathbf{e}^T \boldsymbol{\alpha} \\ & \mathbf{y}^T \boldsymbol{\alpha} = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, \dots, l \end{aligned}$$

in cui  $\mathbf{e} = [1, \dots, 1]^T$  è un vettore composto da numeri 1,  $Q$  è una  $l$  con  $l$  matrice positiva semidefinita,  $Q_{ij} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$ , e  $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$  è la funzione kernel. una volta risolto il problema precedente, utilizzando la relazione primal-dual, l'optimal  $\boldsymbol{\omega}$  soddisfa

$$\boldsymbol{\omega} = \sum_{i=1}^l y_i \alpha_i \phi(\mathbf{x}_i)$$

e la funzione decisionale è

$$\text{sgn}(\boldsymbol{\omega}^T \phi(\mathbf{x}) + b) = \text{sgn}\left(\sum_{i=1}^l y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b\right)$$

salviamo  $y_i \alpha_i \forall i, b$ , nomi dei campi, vettori di supporto ed altre informazioni come parametri kernel nel modello di predizione. [23]

### 2.2.3 K-Nearest Neighbour (K-NN)

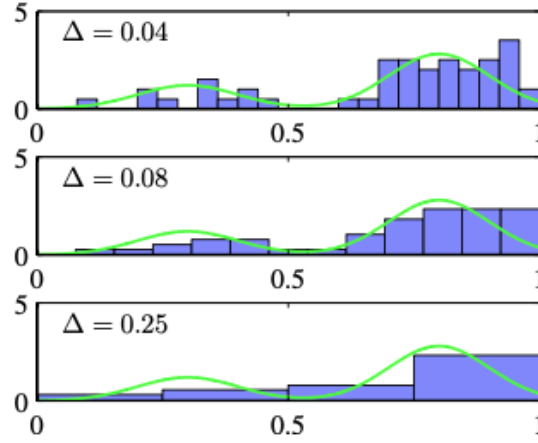
Il K-Nearest Neighbour, o più semplicemente K-NN, si basa invece su un approccio non parametrico. Considerando una singola variabile continua  $x$ , gli istogrammi standard la partizionano in diverse sezioni di spessore  $\Delta_i$  e, successivamente, contano il numero  $n_i$  di osservazioni di  $x$ . Nella figura 2.6 è illustrato l'approccio ad Istogramma per la stima di densità, in cui viene generato un dataset di 50 dati a partire dalla distribuzione sottolineata dalla curva verde.

Supponendo che le stime vengano tracciate da una densità di probabilità sconosciuta  $p(x)$  in uno spazio  $D$ -dimensionale, data una regione  $R$ , la probabilità di massa può essere calcolata come:

$$P = \int_R p(x) dx$$

Supponendo di aver collezionato dati seguendo  $N$  osservazioni, il numero di punti interni a  $R$  saranno distribuiti secondo la distribuzione binomiale

$$\text{Bin}(K|N, P) = \frac{N!}{K!(N-K)!} P^K (1-P)^{1-K}$$



**Figura 2.6:** Histogram Approach

Assumendo che la regione  $R$  sia sufficientemente piccola e che la probabilità  $p(x)$  sia costante su tutta la regione, si ricava che:

$$p(x) = \frac{K}{NV}$$

Partendo dalla precedente formula sono possibili due risoluzioni, si può risolvere  $K$  e determinare il valore di  $V$  dai dati, da cui ha origine la tecnica K-NN, oppure si può risolvere  $V$  e determinare  $K$  dai dati, da cui ha origine l'approccio Kernel. Entrambi i risultati convergeranno alla densità di probabilità del vero per il limite di  $V$  per  $N \rightarrow \infty$ .

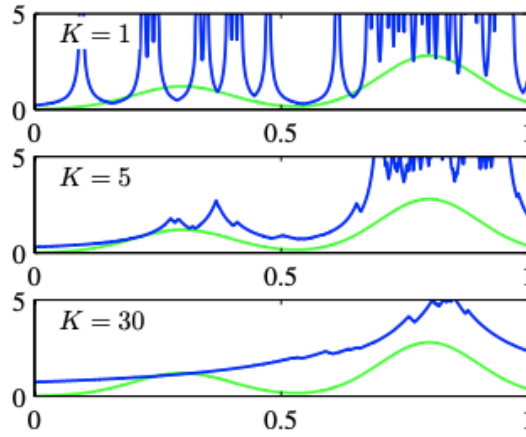
Considerando la seconda, si consideri una piccola sfera centrata nel punto  $x$  su cui stimiamo la densità  $p(x)$  e lasciamo crescere il raggio della sfera fino a che non conterrà interamente  $K$  punti dei dati 2.7.

Nel problema della classificazione, la tecnica della stima di densità K-NN viene applicata ad ogni classe separatamente, e successivamente si utilizza il teorema di Bayes. Supponiamo di avere un dataset composto da  $N_k$  punti in  $C_k$  classi con  $N$  punti totali tali per cui  $\sum_k N_k = N$ . Volendo classificare un nuovo punto  $x$ , disegniamo una sfera centrata in  $x$  contenente precisamente  $K$  punti appartenenti alla stessa classe. Supponendo che la sfera abbia volume  $V$  e che contenga  $N_k$  punti della classe  $C_k$ , provvediamo alla stima della densità associata ad ogni classe

$$p(x|C_k) = \frac{K_k}{N_k V}$$

similmente, la densità non condizionata è data da

$$p(x) = \frac{K}{NV}$$



**Figura 2.7:** Illustrazione dell'estimatore di densità K-NN

mentre la classe è data da

$$p(C_k) = \frac{N_k}{N}$$

possiamo ora combinare le equazioni precedenti utilizzando il teorema di Bayes come anticipato, ottenendo così la probabilità della classe di appartenenza

$$p(C_k|x) = \frac{p(x|C_k)p(C_k)}{p(x)} = \frac{K_k}{K}$$

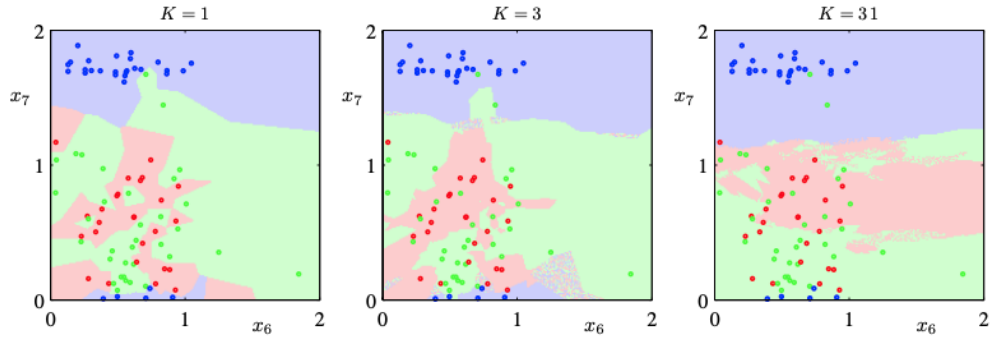
Qualora volessimo minimizzare la probabilità di classificazione erranea, dovremmo assegnare il punto di test  $x$  alla classe con la probabilità più alta, che corrisponde al valore più alto di  $K_k/K$ .

Nella figura 2.8 sono mostrati i risultati dell'applicazione dell'algoritmo K-NN, come previsto,  $K$  controlla il grado di affinamento, così che ad un valore piccolo di  $K$  corrispondano tante piccole regioni per ogni classe, mentre con un valore più grande di  $K$  si ottengono meno regioni ma di dimensioni più estese.

## 2.2.4 Combining Methods

Rispetto ai modelli precedentemente presentati, è stato dimostrato che si possono incrementare le performance utilizzando una combinazione di diversi modelli. Un esempio di questo approccio potrebbe essere selezionare uno dei modelli per effettuare la predizione in cui la scelta del modello è una funzione delle variabili di input, come avviene nei decision tree, che combinati danno origine al Random Forest Classifier. Un altro esempio di applicazione consiste nell'allenare più modelli e successivamente utilizzando la media delle previsioni di ogni modello, questo approccio dà origine al metodo Boosting.

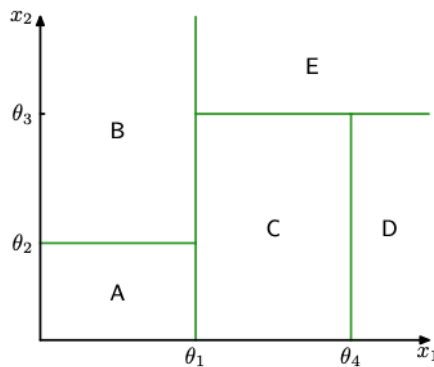




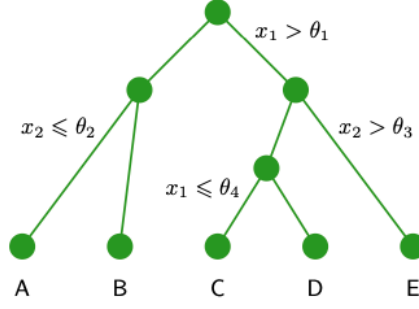
**Figura 2.8:** Grafico con 200 data points che mostra i valori di  $x_6$  raffigurati su  $x_7$

### Random Forest Classifier (RF)

Il Random Forest Classifier, come già anticipato, è un classificatore appartenente alla famiglia dei combining methods che si ottiene dalla combinazione di alberi di decisione (o *decision tree*). Esistono diversi modelli che vengono utilizzati per partizionare lo spazio in input in regioni quadratiche i cui angoli sono allineati con gli assi a cui poi vengono assegnati dei modelli. Questi possono essere descritti come una combinazione di modelli in cui però ad ogni input solo un modello per volta è responsabile della decisione. Nella figura 2.9 è illustrata la suddivisione ricorsiva dello spazio, che può essere tradotta in una struttura ad albero binario rappresentata nella figura 2.10 in cui per ogni nuovo input si determina in quale regione finale verrà posizionato partendo dalla radice sino alle foglie dell'albero. Una proprietà importante dei modelli basati su alberi decisionali che li rende largamente utilizzati nell'ambito delle diagnosi mediche è proprio la possibilità di essere interpretati facilmente poichè costituiti da decisioni binarie applicate a singoli input.



**Figura 2.9:** Spazio di input a due dimensioni partizionato. adapted from [22]



**Figura 2.10:** Binary Tree. adapted from [22]

Scendendo nello specifico, come prima cosa è necessario approfondire l'algoritmo alla base del Random Forest [24], dato un vettore random  $p$ -dimensionale  $X = (X_1, \dots, X_p)^T$  che rappresenta i valori di input, assumiamo una funzione di distribuzione congiunta non nota  $P_{XY}(X, Y)$ . L'obiettivo principale è trovare una funzione di predizione  $f(X)$  per predire  $Y$ . La funzione di predizione è determinata dalla funzione di perdita  $L(Y, f(X))$  ed è definita per minimizzare il valore atteso della perdita

$$E_{XY}(L(Y, f(X)))$$

La scelta della funzione  $L$ , per i problemi di classificazione, tipicamente ricade sulla funzione *zero-uno*

$$L(Y, f(X)) = I(Y \neq f(X)) = \begin{cases} 0 & \text{if } Y = f(X) \\ 1 & \text{otherwise} \end{cases} \quad (2.1)$$

Se il set dei valori possibili di  $Y$  viene indicato con  $\gamma$ , minimizzando  $E_{XY}(L(Y, f(X)))$  per la funzione gradino unitaria precedentemente illustrata si ottiene

$$f(x) = \arg \max_{y \in \gamma} P(Y = y | X = x)$$

anche conosciuta come Regola di Bayes. L'Ensemble determina  $f$  al fine di costituire un insieme chiamato *Base Learners*  $h_1(x), \dots, h_J(x)$ , i cui componenti vengono combinati per poter creare il predittore combinato. Nello specifico  $f(x)$  è la classe predetta più frequentemente

$$f(x) = \arg \max_{y \in \gamma} \sum_{j=1}^J I(y = h_j(x))$$

Come già anticipato all'inizio della sezione corrente, gli alberi utilizzati nel Random Forest si basano su alberi binari ricorsivi la cui radice comprende l'intero spazio di predizione, ad ogni nodo (ad eccezione per i nodi foglia) corrisponde la divisione in

due nodi discendenti. Una variabile predittiva categorica  $X_i$  prende i valori da un insieme finito di categorie  $S_i = \{s_{i,1}, \dots, s_{i,m}\}$ . La divisione in due nodi porta ad un sottoinsieme delle precedenti categorie  $S \subset S_i$  a sinistra e le restanti categorie non presenti nel primo insieme andranno al nodo di destra. Le condizioni di divisione ai nodi di un albero vengono scelte considerando tutte le divisioni possibili in quel nodo e scegliendo la migliore a seconda dei criteri scelti. Nel contesto della classificazione, in cui si hanno  $K$  classi denotate  $\dots, K$  un criterio di divisione tipico è il Gini index:

$$Q = \sum_{k \neq k'}^K \hat{p}_k \hat{p}_{k'}$$

in cui  $\hat{p}_k$  è la percentuale di osservazioni della classe  $k$  nel nodo:

$$\hat{p}_k = \frac{1}{n} \sum_{i=1}^n I(y_i = k)$$

il criterio di divisione fornisce una misura di "purezza" per un nodo in cui grandi valori corrispondono ad una classificazione impura. Una volta che il criterio di divisione è stato scelto e i dati partizionati in due nodi discendenti, ogni nodo viene trattato allo stesso modo del nodo di origine. La procedura prosegue ricorsivamente sino a quanto non si raggiunge un criterio di terminazione. Si ottiene un valore predetto calcolando la classe più frequente ottenuta.

### Gradient Boosting Classifier (GB)

Nelle Gradient Boosting Machines [25], il processo di apprendimento si basa sull'utilizzo di nuovi modelli al fine di fornire una stima migliore dei risultati. La struttura di base di questo modello prevede la costruzione di un nuovo classificatore base tenendo come obiettivo la massimizzazione della correlazione con il gradiente negativo della funzione di perdita. Quest'ultima può essere scelta dal programmatore avendo la possibilità di implementarne una a partire da richieste specifiche. Considerando il problema della stima della funzione, trovandoci in un approccio di Supervised Learning, i dati vengono forniti con un set di labels, quindi partendo da un dataset  $x = (x_1, \dots, x_d)$ , l'obiettivo iniziale sarà ricostruire una funzione di dipendenza  $x \xrightarrow{f} y$  a partire dai dati, stimando la funzione  $\hat{f}(x)$  affinché la funzione di perdita  $\Psi(y, f)$  venga minimizzata:

$$\hat{f}(x) = y$$

$$\hat{f}(x) = \arg \min_{f(x)} \Psi(y, f(x))$$

al fine di rendere trattabile la funzione di stima, si può restringere la funzione di ricerca dello spazio in un insieme parametrico di funzioni  $\hat{f}(x, \hat{\theta})$ , ottenendo così il

seguente problema

$$\hat{f}(x) = f(x, \hat{\theta})$$

in cui riscrivendo il problema in termini di valori attesi il risultato sarebbe equivalente alla minimizzazione del valore atteso della funzione di perdita

$$\hat{\theta} = \arg \min_{\theta} E_x[E_y(\Psi[y, f(x, \theta)])|x]$$

una volta affrontato il problema della funzione di stima, occorre occuparsi dell'ottimizzazione numerica, quindi, date  $M$  iterazioni, il parametro stimato può essere riscritto come

$$\hat{\theta} = \sum_{i=1}^M \hat{\theta}_i$$

applicandolo alla funzione di perdita si ottiene così la funzione  $J(\theta)$  che, dati  $N$  punti  $(x, y)_{i=1}^N$ , si vuole diminuire la sua funzione di perdita in base ai dati osservati

$$J(\theta) = \sum_{i=1}^N \Psi(y_i, f(x_i, \hat{\theta}))$$

La procedura classica di ottimizzazione si basa su miglioramenti consecutivi basati sul gradiente della funzione di perdita  $\nabla J(\theta)$ . In seguito sono illustrati i passi per la procedura di ottimizzazione

1. Inizializzare del parametro stimato  $\hat{\theta}_0$ , per ogni iterazione  $t$  ripetere:
2. Ricavare un parametro stimato dalle iterazioni precedenti:

$$\hat{\theta}^t = \sum_{i=0}^{t-1} \hat{\theta}_i$$

3. Calcolare il gradiente della funzione di perdita  $\nabla J(\theta)$  a partire dal parametro di stimato dell'insieme:

$$\nabla J(\theta) = \{\nabla J(\theta_i)\} = \left[ \frac{\partial J(\theta)}{\partial J(\theta_i)} \right]_{\theta=\hat{\theta}^t}$$

4. Calcolare il nuovo parametro stimato  $\hat{\theta}_t$

$$\hat{\theta}_t = -\nabla J(\theta)$$

5. Aggiungere il nuovo valore stimato  $\hat{\theta}_t$  all'insieme.

Ciò che distingue i metodi convenzionali di machine learning dai metodi di Boosting, è proprio lo svolgimento dell'ottimizzazione all'interno dello spazio delle funzioni. La funzione di stima  $\hat{f}$  viene parametrizzata in nel metodo sottoillustrato

$$\hat{f}(x) = \hat{f}^M(x) = \sum_{i=0}^M \hat{f}_i(x)$$

in cui M rappresenta il numero di iterazioni,  $\hat{f}_0$  è la funzione di partenza mentre  $\{\hat{f}_i\}_{i=1}^M$  sono le funzioni calcolate, anche chiamate "boots". Per rendere applicabile alla realtà questo approccio, si può utilizzare l'approccio spiegato in seguito che prevede l'utilizzo di funzioni parametrizzate "base-learner"  $h(x, \theta)$ , con le quali si applica la regola di ottimizzazione seguente

$$\begin{aligned} \hat{f} &= \hat{f}_{t-1} + \rho_t h(x, \theta_t) \\ (\rho_t, \theta_t) &= \arg \min_{\rho, \theta} \sum_{i=1}^N \Psi(y_i, \hat{f}_{t-1}) + \rho h(x_i, \theta) \end{aligned}$$

Avendo spiegato tutti i passaggi e le componenti fondamentali si può ora affrontare la struttura dell'algoritmo Gradient Boosting, come prima cosa è bene sottolineare che data una funzione di perdita  $\Psi(y, f)$  e/o un base-learner  $h(x, \theta)$ , la soluzione dei parametri potrebbe essere difficile da ottenere, così si è proposto di scegliere una nuova funzione  $h(x, \theta_t)$  che sia parallela al gradiente negativo  $\{g_t(x_i)\}_{i=1}^N$  sui dati osservati

$$g_t(x) = E_y \left[ \frac{\partial \Psi(y, f(x))}{\partial f(x)} \middle| x \right]_{f(x)=\hat{f}^{t-1}(x)}$$

in questo modo, al posto di cercare la soluzione generale nello spazio delle funzioni, si può scegliere una nuova funzione che sia maggiormente correlata con  $-g_t(x)$ , questo permette di semplificare il problema di ottimizzazione con il metodo dei quadrati minimi:

$$(\rho_t, \theta_t) = \arg \min_{\rho, \theta} \sum_{i=1}^N [-g_t(x_i) + \rho h(x_i, \theta)]^2$$

## 2.3 Elementi di Statistica e terminologia per la creazione e validazione dei modelli

In questa sezione verranno riassunti i concetti chiave di statistica e di machine learning [26] al fine di poter comprendere al meglio gli argomenti che verranno trattati nel capitolo successivo.

### 2.3.1 Statistica

Nel seguente elenco saranno riassunti solo i principali concetti di statistica necessari alla comprensione della realizzazione delle features scelte per il modello di machine learning.

- **Media:** è la semplice media aritmetica, calcolata sommando i valori e dividendo il risultato con il numero di valori precedentemente sommati
- **Varianza:** la media degli scarti quadratici rispetto alla media

$$varianza = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

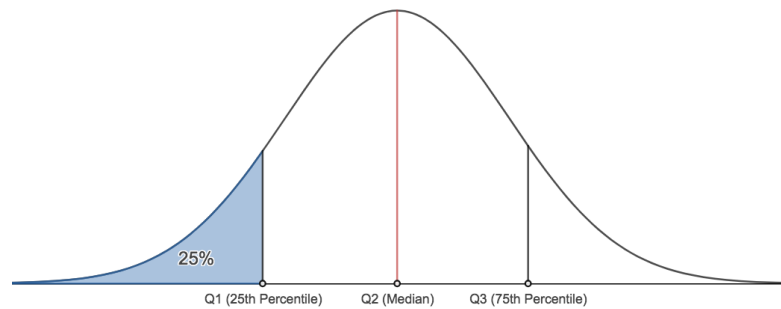
- **Deviazione standard:** è la radice quadrata della varianza, applicando la radice quadrata si può calcolare la dispersione rispetto alla variabile originale

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

- **Disribuzione Normale:** La popolazione di tutti i campioni possibili di dimensione  $n$  da una popolazione con media  $\mu$  e varianza  $\sigma^2$  può essere rappresentata da una distribuzione normale:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad X \sim N(\mu, \sigma^2)$$

- **Quantili:** un quantile è un valore che divide la popolazione in due parti. I percentili indicano la percentuale di dati sotto al valore. Nella figura 2.11 sono rappresentati il 25° e 75° percentile



**Figura 2.11:** 25° e 75° Percentile in una distribuzione Normale

- **Correlazione di Pearson:** Esprime la relazione di linearità tra due variabili statistiche

$$\rho = \frac{\sum (x_i - x)(y_i - y)}{\sqrt{\sum (x_i - x)^2 \sum (y_i - y)^2}}$$

### 2.3.2 Matrice di confusione

		Actual Values	
		Positive(1)	Negative(0)
Predicted Values	Positive(1)	<b>TP</b>	<b>FP</b>
	Negative(0)	<b>FN</b>	<b>TN</b>

Figura 2.12: Matrice di confusione

- **True positives (TPs):** Casi in cui viene predetta l'appartenenza ad una classe corretta.
- **True negatives (TNs):** Casi in cui viene predetta correttamente la non appartenenza ad una classe.
- **False positives (FPs):** Casi in cui vien predetta l'appartenenza ad una classe quando invece non è così.
- **False negatives (FNs):** Casi in cui viene predetta la non appartenenza ad una classe quando non dovrebbe
- **Accuracy:** frazione delle predizioni corrette

$$\frac{TP + TN}{TP + FP + TN + FN}$$

- **Precision (P)**: Descrive la frazione di casi in cui la predizione è corretta

$$\frac{TP}{TP + FP}$$

- **Recall (R)/sensitivity/true positive rate**: tra le classificazioni corrette, qual è la frazione di appartenenti alla classe

$$\frac{TP}{TP + FN}$$

- **F1 score (F1)**: Media armonica di *precision* e *recall*

$$F_1 = \frac{2}{\frac{1}{P} + \frac{1}{R}} \gg F_1 = \frac{2PR}{P + R}$$

- **Specificity**: Frazione predetta come non appartenente alla classe, ossia frazione dei falsi positivi

$$\frac{TN}{TN + FP}$$

- **Area under curve (ROC)**: Grafico ottenuto tra il punteggio dei positivi (TPR) e i falsi positivi (FPR)



## Capitolo 3

# Metodi e materiali utilizzati

### 3.1 Questionari

Spesso la ricerca in ambito medico fa affidamento su dati ottenuti tramite la somministrazione di questionari al fine di determinare lo stato di salute dei pazienti. Tuttavia la soggettività delle risposte è uno dei potenziali problemi che limita l'accuratezza dei dati raccolti tramite questionari. Un metodo utile nella validazione dei questionari è eseguire analisi quantitative su ciò che si sta studiando al fine di poter mettere a confronto l'analisi qualitativa con quella quantitativa. I questionari che sono stati utilizzati in questo lavoro sono stati entrambi validati e pubblicati.

#### 3.1.1 sPSQI

L'acronimo sPSQI[27] sta per short Pittsburgh Sleep Quality Index, è appunto la versione ridotta del preesistente PSQI pubblicato nel 1988 al fine di dare una misurazione qualitativa data da un'autovalutazione della qualità del sonno attraverso appunto la somministrazione di un questionario. Si è giunti alla riduzione del precedente questionario nel 2018 attraverso l'articolo pubblicato da Oluremi A. Famodu, la riduzione del numero di domande è stata effettuata poiché proprio la lunghezza dei questionari risultava essere uno dei problemi che limitavano l'accuratezza dei risultati. Sono stati effettuati dei test statistici prima per valutare quali domande selezionare, successivamente per determinare la validità del questionario ridotto. Nella tabella 3.1 sono rappresentate le tredici domande selezionate che costituiscono l'sPSQI.

---

**Domande**

---

1. Durante l'ultimo mese a che ora sei andato a dormire?
  2. Durante l'ultimo mese, quanti minuti hai impiegato per addormentarti?
  3. Durante l'ultimo mese a che ora ti sei svegliato?
  4. Durante l'ultimo mese, quante ore di sonno hai avuto ogni notte?
  5. a. Durante l'ultimo mese, quante volte hai riscontrato difficoltà nel sonno dovute a: Non riuscire ad addormentarsi entro 30 minuti?
  - 5.b. Durante l'ultimo mese, quante volte hai riscontrato difficoltà nel sonno dovute a: Svegliarsi durante la notte o al mattino presto?
  - 5.c. Durante l'ultimo mese, quante volte hai riscontrato difficoltà nel sonno dovute a: Difficoltà a respirare?
  - 5.d. Durante l'ultimo mese, quante volte hai riscontrato difficoltà nel sonno dovute a: Tosse o russamento?
  - 5.e. Durante l'ultimo mese, quante volte hai riscontrato difficoltà nel sonno dovute a: Sentire caldo?
  - 5.f. Durante l'ultimo mese, quante volte hai riscontrato difficoltà nel sonno dovute a: Brutti sogni?
  - 5.g. Durante l'ultimo mese, quante volte hai riscontrato difficoltà nel sonno dovute a: dolori nel sonno?
  6. Nell'ultimo mese quante volte hai riscontrato difficoltà nello stare sveglio alla guida, durante i pasti, o durante momenti di compagnia?
  7. Nell'ultimo mese quante volte hai riscontrato difficoltà nel mantenere l'entusiasmo nel fare le cose?
- 

**Tabella 3.1:** Domande sPSQI

Nella tabella 3.2 è illustrato il metodo con cui sono stati calcolati i punteggi relativi allo scoring del questionario sPSQI. Dopo aver calcolato tutte le componenti (illustrate nella colonna sinistra della tabella sopracitata), queste vengono sommate ottenendo così un punteggio. Secondo il Global PSQI un punteggio maggiore di 4 è indice di una qualità del sonno scarsa.

Componente	calcolo punteggio
<i>Component 1: sleep latency</i>	(2) Score (<15 (0), 16–30 min (1), 31–60 min (2), >60 min (3)) + (5a) Score (if sum is equal 0 = 0; 1-2 = 1; 3-4 = 2; 5-6 = 3)
<i>Component 2: sleep duration</i>	(4) Score (>7 (0), 6-7 (1), 5-6 (2), <5 (3))
<i>Component 3: sleep disturbances</i>	Examine questions (5b) to (5g) and assign scores for each questions as follows Response: Score Not during the past month: 0. Less than once a week: 1. Once or twice a week: 2. Three or more times a week: 3. sum of scores (5b) to (5g) (0=0, 1=1; >6 12 = 2; >12 = 3)
<i>Component 5: daytime dysfunction</i>	(6)+(7) Score (0=0; 1-2=1; 3-4=2; 5-6=3)

Tabella 3.2: Calcolo score sPSQI

### 3.1.2 SLEEPS

I partecipanti hanno compilato un secondo questionario, SLEEPS [28], SLEEP Survey, costituito da 21 domande che permettono di dare un'interpretazione soggettiva al sonno-veglia dei soggetti in analisi. L'obiettivo principale di questo questionario è proprio determinare la relazione tra le abitudini dei soggetti e la loro qualità del sonno. Le domande che lo costituiscono sono raccolte nella tabella 3.3.

---

**Domande**


---

1. Indicare se si ha avuto una diagnosi di COVID-19
  2. Indicare se si è affetti da OSAS (sindrome delle apnee ostruttive del sonno)
  3. Indicare se si soffre di insonnia
  4. Indicare se si è attualmente studenti presso università
  5. Indicare se attualmente si sta lavorando
  6. Indicare se si seguono lezioni a distanza/ lavoro in modalità da remoto
  8. Fino a che ora utilizzi strumenti elettronici per studio/lavoro?
  9. Utilizzi un filtro per la luce blu?
  10. Quanto tempo trascorri fuori casa nei giorni feriali (media giornaliera)?
  11. Quanto tempo trascorri fuori casa nel fine settimana (media giornaliera)?
  12. Quanto tempo dedichi settimanalmente all'attività fisica all'aria aperta?
  13. Quanto tempo dedichi settimanalmente all'attività fisica in casa o in palestra?
  14. Quanto tempo dedichi settimanalmente ai tuoi hobby?
  15. Ti capita di svegliarti durante la notte?
  16. Ti capita di alzarti dal letto la notte?
  17. Ti capita di avere difficoltà ad alzarti?
  18. Ti capita di svegliarti sentendoti stanco?
  19. Ti capita di sentirti stanco e affaticato durante il giorno?
  20. Ti capita di sentirti stanco e affaticato alla sera ma avere difficoltà ad addormentarti?
  21. Su una scala da 0 a 4, come definiresti la qualità del tuo sonno?
- 

**Tabella 3.3:** Domande SLEEPS

## 3.2 Sensori inerziali

I disturbi come Acinesie e Ipocinesie sono spesso sottostimati e sono pochi gli studi presenti in letteratura che si soffermano sullo studio di sintomi notturni nei pazienti con PD in relazione alle funzioni motorie e alla qualità del sonno. [29] L'obiettivo principale di questo studio è stato infatti quello di effettuare un'analisi quantitativa dei disturbi nel sonno al fine di poter successivamente avere un dataset su cui effettuare un'analisi più approfondita. Per rendere possibile questa raccolta dati sono stati utilizzati dei sensori inerziali indossati durante la notte.

### 3.2.1 Shimmer3

Il sensore che è stato utilizzato in questa ricerca è lo Shimmer3 raffigurato in figura 3.1. Le specifiche Hardware del sensore in questione, sono riassunte nella figura 3.2,



**Figura 3.1:** Sensore Shimmer3

in cui è presente un diagramma a blocchi della scheda principale con i componenti e le interconnessioni tra i dispositivi integrati. L'elemento centrale è il processore low-power MSP430F5437A che controlla le operazioni del device. La CPU configura e controlla diverse periferiche attraverso i pin I/O, alcuni dei quali disponibili attraverso il connettore di espansione interna/esterna. La CPU ha a disposizione anche un convertitore analogico-digitale (ADC) integrato a 16 canali e 12 bit, utilizzato per identificare il rumore di segnale dato dall'accelerometro, batteria o dai sensori di espansione che possono essere collegati. L'espansione esterna offre anche la possibilità di entrare in comunicazione con la scheda principale attraverso la docking station. Il sensore ha a disposizione anche uno slot per microSD per uno storage aggiuntivo. È disponibile anche la connettività bluetooth per uno streaming dei dati wireless.

La configurazione impostata per la raccolta dati è stata la seguente:

- Accelerometer (x, y, z): **ON**
- Gyroscope (x, y, z): **ON**
- Magnetometer (x, y, z): **ON**
- Sampling rate: **128 Hz**

Nella figura 3.3 è raffigurato il layout della scheda principale montata nello Shimmer3 con ogni componente etichettato.

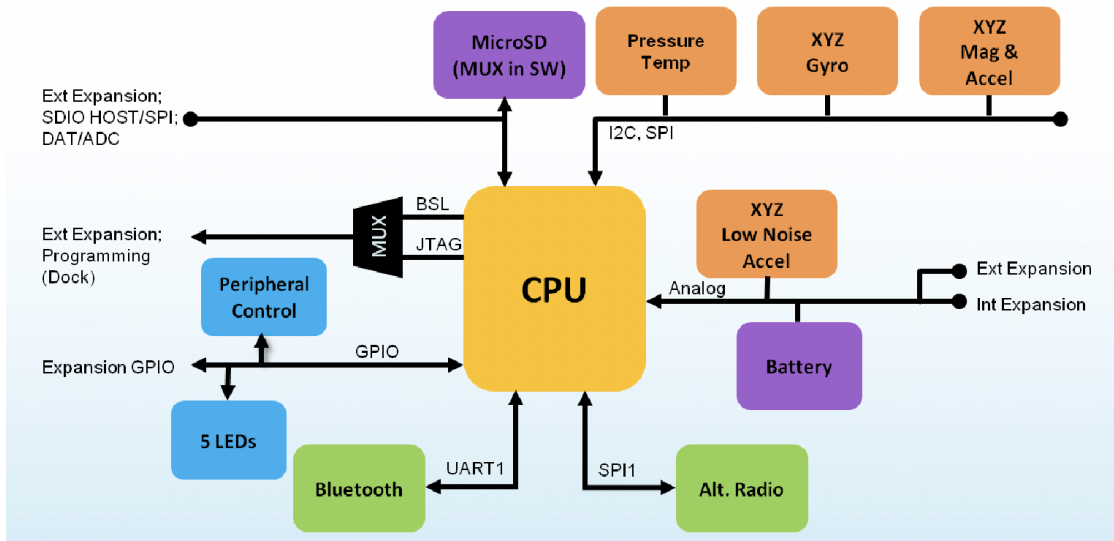


Figura 3.2: Shimmer3 core component diagram

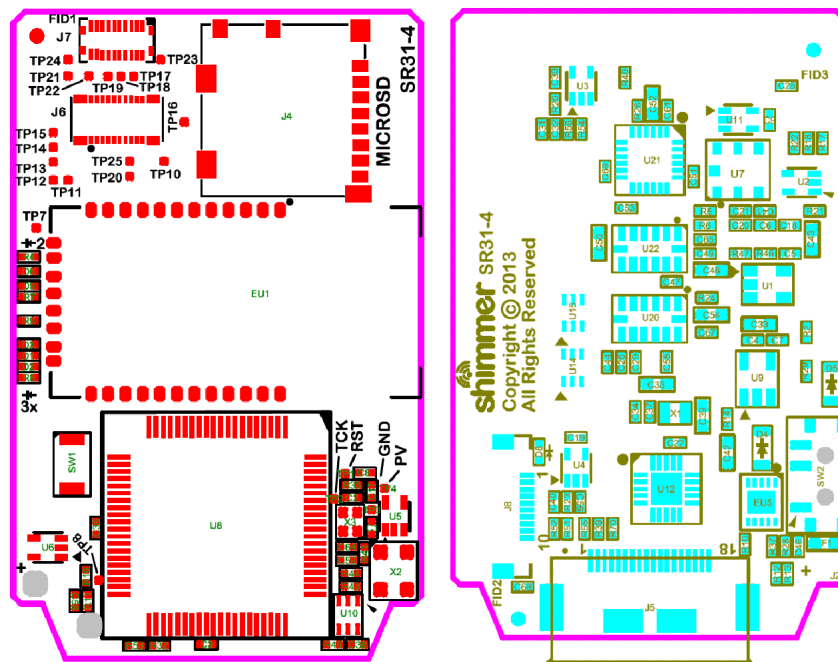


Figura 3.3: Shimmer3 mainboard layout (a sinistra vista dall'alto, a destra dal basso)

## 3.3 Raccolta dati

### 3.3.1 Posizionamento Shimmer3

Nei soggetti che si sono prestati alla raccolta dati, sono state effettuate delle misurazioni notturne grazie all'utilizzo dei sensori inerziali Shimmer3 di cui abbiamo già parlato nel paragrafo precedente. Nello specifico i sensori sono stati indossati durante la notte, posizionandoli in prossimità dello sterno attraverso una fascia elastica legata intorno al torace. Il posizionamento sui soggetti è rappresentato nella figura 3.4. Grazie al posizionamento scelto è stato possibile registrare ogni movimento di rotazione del busto. Per ogni soggetto è stata effettuata una registrazione di una sola notte.

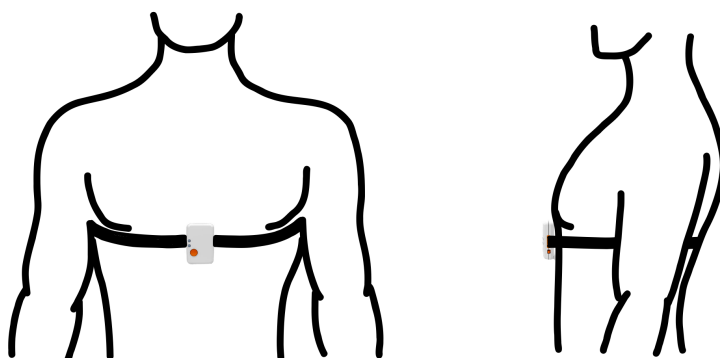
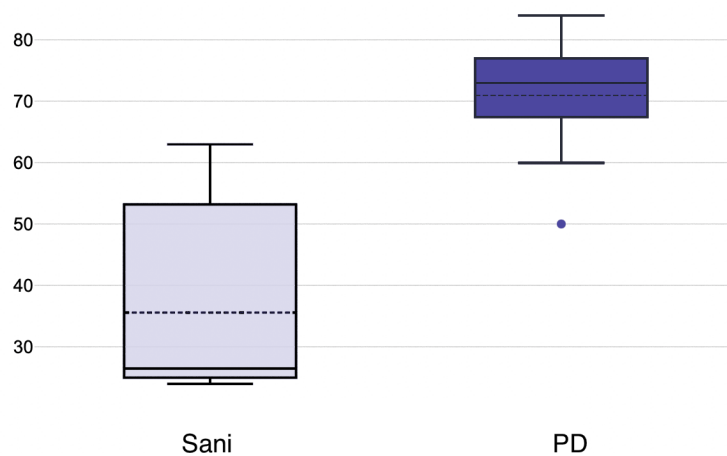


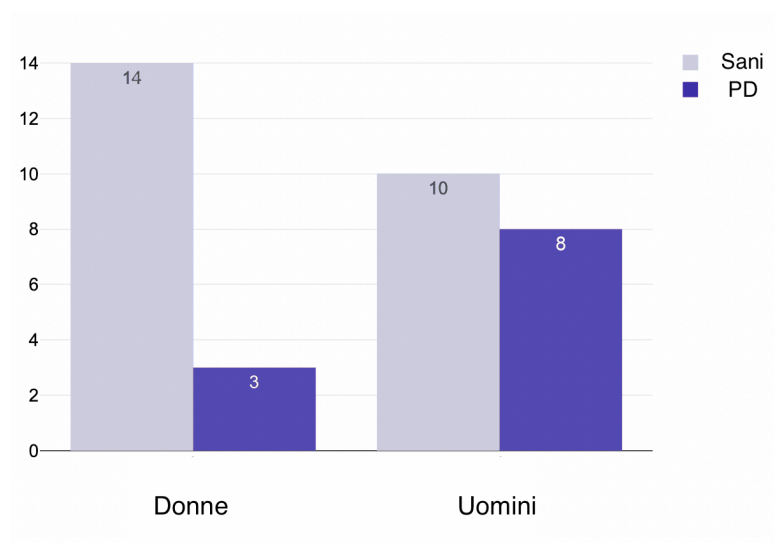
Figura 3.4: Shimmer3 indossato

### 3.3.2 Campione rappresentativo

Trentadue soggetti, tra cui pazienti con diagnosi di Parkinson e controlli, hanno preso parte alla sperimentazione. Gli otto pazienti con diagnosi di Parkinson sono stati contattati grazie al Centro di Medicina del Sonno dell'Ospedale Molinette di Torino, all'Associazione Amici Parkinsoniani Piemonte e infine grazie alla Residenza Protetta Madre Rubatto di Loano (SV). In seguito sono riportate le statistiche riguardanti Età (con divisione Sani/PD) in figura 3.5, Sesso (con divisione Sani/PD) in figura 3.6



**Figura 3.5:** BoxPlot con distribuzione dell'età nei soggetti sani e soggetti con diagnosi di Parkinson



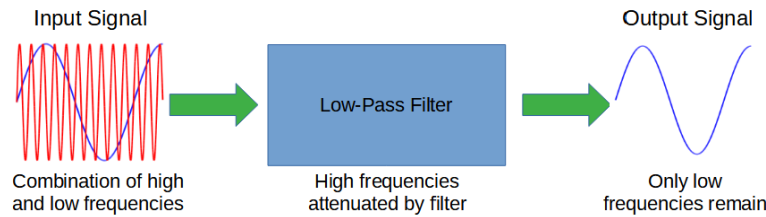
**Figura 3.6:** Istogramma rappresentante il numero di soggetti e il corrispondente sesso, suddiviso tra soggetti sani e soggetti con diagnosi di Parkinson



## 3.4 Elaborazione dati

### 3.4.1 Prefiltraggio

Dopo aver importato i dati, un passaggio necessario e fondamentale è stato quello di eliminare eventuale rumore dato da disturbi causati da problemi di utilizzo o limitazioni hardware dei sensori utilizzati. Su tutti i dati raccolti sono stati applicati filtri passabasso. Il funzionamento del filtro passabasso può essere riassunto dal diagramma in figura 3.7 in cui un segnale che presenta disturbi, grazie all'applicazione di un filtro vengono eliminati i disturbi, ottenendo così un segnale più stabile e *pulito*.



**Figura 3.7:** Funzionamento filtro passabasso

Il lavoro di estrazione delle features è stato effettuato a partire dai risultati di questo prefiltraggio, nelle figure successive si possono osservare i grafici di Accelerometro (figura 3.8), Giroscopio (figura 3.9) e Magnetometro (figura 3.10), filtrati e messi a confronto tra soggetti sani e soggetti con diagnosi di Parkinson. Una stima qualitativa può già essere effettuata osservando i grafici dei tre sensori raccolti messi a confronto tra Sani e PD, dalla figura 3.8 si può notare come l'attività motoria notturna del soggetto sano sia nettamente superiore rispetto a quella del soggetto PD. Proseguendo con la figura 3.9 si può notare come l'ampiezza dei segnali, e dunque i valori di Accelerazione Angolare, sia maggiore nei soggetti sani rispetto ai soggetti PD, che in questo caso non solo mostrano una minore attività, ma anche i movimenti registrati mostrano una piccola variazione di accelerazione angolare. Nella figura 3.10, allo stesso modo dei grafici dell'accelerometro, si può osservare che le variazioni di campo magnetico, e quindi il movimento in termini di posizione assoluta, siano molto più frequenti nel soggetto sano rispetto al soggetto PD.

### 3.4.2 Estrazione delle Features

Al fine di poter rendere efficace la classificazione sono state scelte le seguenti features basandosi sui principali parametri osservati negli studi sui disturbi del sonno [15]. Buona parte delle features corrispondono ad un calcolo effettuato su

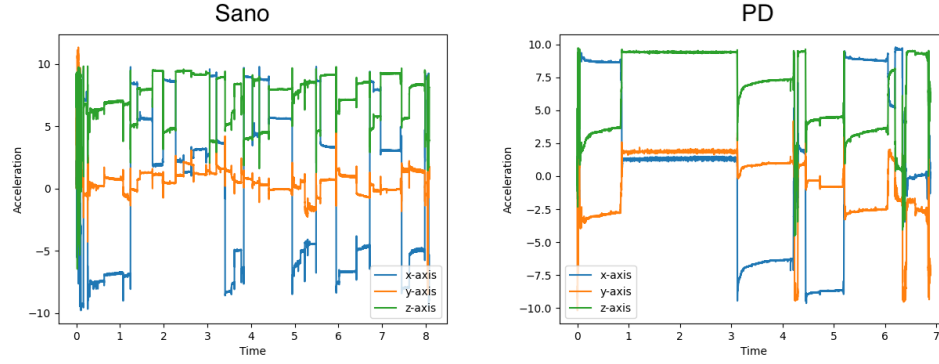


Figura 3.8: Accelerometro

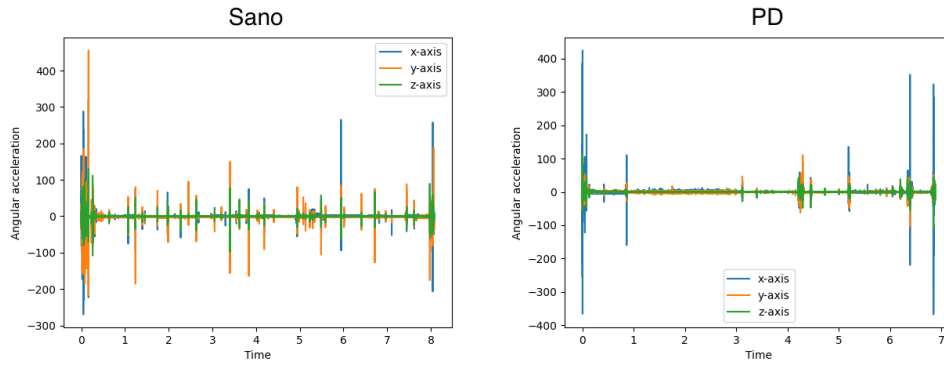
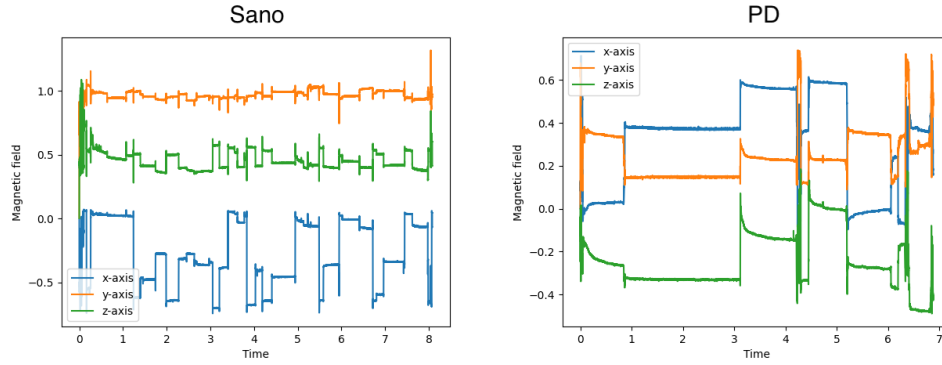


Figura 3.9: Giroscopio

ogni *evento notturno*, in questo caso possono essere considerati *eventi notturni* azioni quali movimenti lungo uno o più assi, per esempio rotazioni del busto, o cambiamenti di posizione da supino a seduto. In alcune features, considerando la presenza di più movimenti nell'arco della raccolta dati notturna, i valori sono stati inseriti in array dai quali sono state calcolate media, deviazione standard, 25° percentile, 75° percentile e simmetria statistica. Gli eventi in questione sono stati identificati temporalmente in corrispondenza dei punti in cui il giroscopio registra dei picchi, successivamente i calcoli sono stati effettuati a partire dai dati provenienti da accelerometro, giroscopio e magnetometro. In seguito sono stati elencate le features sopracitate con i relativi pseudocodici realizzati.

- **tot\_hrs**: Durata totale del sonno, registrata in ore.

$$\frac{\text{lunghezza}(\text{dati\_raccolti})}{\text{frequenza} \cdot 3600}$$



**Figura 3.10:** Magnetometro

- **tilt\_angle**: Angolo tra asse di gravità e asse longitudinale.

$$\left| \arctan \left( \frac{\text{mean}(\text{acceleration\_z})}{\text{mean}(\text{acceleration\_y})} \right) \cdot \frac{180}{\pi} \right|$$

- **number\_turns**: Numero di giri registrati nel corso di tutta la notte.

$$\text{lunghezza}(\text{eventi\_rotazione})$$

- **turning\_interval**: Intervallo di tempo intercorso tra un giro e il successivo.

$$\frac{\text{differenza\_discreta}(\text{eventi\_rotazione})}{fs \cdot 60}$$

- **turning\_ang\_vel**: Velocità angolare di ogni movimento, data dall'altezza del picco dato dal giroscopio, in corrispondenza del quale si verifica il movimento.
- **turning\_ang\_acc**: Accelerazione angolare di ogni movimento, misurata dal giroscopio.
- **turning\_duration**: Durata di ogni movimento, data dalla larghezza del picco dato dal giroscopio, in corrispondenza del quale si verifica il movimento.
- **turning\_mag**: Variazione del campo magnetico durante ogni movimento
- **AI**: Indice di attività basato sulle epoche.

```

1  rumore_shimmer = 27.5e-3 #rumore Shimmer3
2  N = lunghezza(dati_accelerometro)/numero_samples
3  for i in N

```

```

4         per x, y e z
5             xi = dati_accelerometro(da numero_samples*(i-1)
6                                     a i*numero_samples)
7         per x, y e z
8             var_x = (deviazione_standard(xi)-rumore_shimmer)^2
9

```

- **M1**: Media mobile dell'indice di attività precedente, calcolata con una finestra di 1 minuto.
- **M2**: Media mobile dell'indice di attività precedente, calcolata con una finestra di 2 minuti.
- **sPSQI**: Punteggio calcolato a seguito della somministrazione del questionario sPSQI.
- **SLEEPS**: Punteggio calcolato a seguito della somministrazione del questionario SLEEPS.

### 3.4.3 Feature selection

Una volta generate le features, è stato necessario effettuare una feature selection per permettere una classificazione ottimale. In questo caso una volta elaborati i dati di tutti i soggetti partecipanti allo studio, si è calcolato il Coefficiente di Correlazione di Pearson [30]. Scendendo nello specifico, sono state messe a confronto le varie features con la feature principale, ossia quella inerente alla diagnosi di Parkinson (registrata come valore binario 0/1), nel caso della classificazione binaria Parkinson vs Non-Parkinson, e alla nuova feature "GoodSleeper" proveniente dall'esito dei questionari di autodiagnosi sPSQI e SLEEPS (anche in questo caso, registrata come valore binario 0/1). La variabile principale e la variabile riferita alla feature da selezionare o meno per semplicità verranno indicate rispettivamente con  $a$  e  $b$ . il Coefficiente di correlazione di Pearson è definito come

$$\rho(a, b) = \frac{W(ab)}{\sigma_a \sigma_b}$$

in cui  $E(ab)$  corrisponde alla cross-correlazione tra  $a$  e  $b$ , mentre  $\sigma_a^2 = E(a^2)$  e  $\sigma_b^2 = E(b^2)$  sono rispettivamente le varianze delle variabili  $a$  e  $b$ .

## 3.5 Classificazione

### 3.5.1 Classificazione binaria: Parkinson vs Non-Parkinson

Per quanto riguarda la Classificazione Binaria basata sull'identificazione di soggetti con diagnosi di Parkinson verso soggetti Non-Parkinson, una volta raccolti i dati

della motilità notturna, per ogni soggetto sono state calcolate le features inerenti al sonno registrato e sono poi stati calcolati i punteggi dei questionari attenendosi ai criteri di valutazione riportati nella tabella 3.2. Dopo aver effettuato la feature selection, si sono scelti quattro classificatori con i quali si sarebbe effettuata la classificazione, ossia:

- Support Vector Classifier
- K-Nearest Neighbour Classifier
- Random Forest Classifier
- Gradient Boosting Classifier

Una volta scelti i classificatori da mettere a confronto in termini di prestazioni, si è poi scelto di fare un'ottimizzazione degli iperparametri, basandoci sull'F1-score, questo poiché l'F1-Score offre una misura di accuratezza dei test effettuati tenendo conto di precisione e recupero dei test e offre capacità di generalizzazione più ampia nel caso di dataset sbilanciati. Ottenuti gli iperparametri ottimali per una classificazione che permettesse una resa massima in termini di F1-Score, è stata effettuata la classificazione finale con il confronto tra i classificatori per la scelta del classificatore migliore.

### 3.5.2 Classificazione Good/Bad Sleepers

Partendo dai dati ottenuti dai questionari sPSQI e SLEEPS approfonditi nel capitolo precedente, si è scelto di implementare una classificazione dalla quale ottenere una effettiva correlazione tra percezione soggettiva del sonno e dati oggettivi. I dati e le features utilizzati per la classificazione Sani/PD sono stati riutilizzati messi in correlazione con i risultati ottenuti dai questionari grazie ai quali si è attribuito il valore 1 per i Good Sleepers, e 0 per i Bad Sleepers. Dopo aver effettuato la feature selection, si sono scelti gli stessi classificatori con i quali effettuare la classificazione, ossia:

- Support Vector Classifier
- K-Nearest Neighbour Classifier
- Random Forest Classifier
- Gradient Boosting Classifier

Anche in questo caso si è scelto di eseguire un'ottimizzazione basata sull'F1-score. A seguito dell'ottimizzazione degli iperparametri sono stati implementati i classificatori e confrontati i valori degli stessi al fine di determinarne il migliore.

## Capitolo 4

# Risultati

### 4.1 Classificazione Sani/PD

#### 4.1.1 Feature Selection

Dopo aver generato le features, queste sono state a loro volta selezionate basandosi sul Coefficiente di Correlazione di Pearson in relazione alla feature principale, ossia quella riportante la presenza o meno di Diagnosi di Parkinson. Le features che si è scelto di mantenere riportavano un valore del Coefficiente di Correlazione di Pearson strettamente maggiore a 0.3, nella tabella 4.1 sono riportati tutti i valori di correlazione, tra questi sono stati evidenziati i valori che si trovano al di sotto della soglia stabilita che hanno comportato la rimozione della corrispondente feature dalla classificazione.

Feature	Correlazione
tot_hrs	0.244529
mean_tilt_angle	0.407095
std_tilt_angle	0.109419
pc25_tilt_angle	0.190859
pc75_tilt_angle	0.720037
skew_tilt_angle	0.181808
number_turns	0.118904
mean_turning_interval	0.150986
std_turning_interval	0.150562
pc25_turning_interval	0.278792
pc75_turning_interval	0.136212
skew_turning_interval	0.151895
mean_turning_ang_vel	0.569823
std_turning_ang_vel	0.075118
pc25_turning_ang_vel	0.547756
pc75_turning_ang_vel	0.719577
skew_turning_ang_vel	0.426163
mean_turning_duration	0.082138
std_turning_duration	0.075766
pc25_turning_duration	0.184469
pc75_turning_duration	0.056480
skew_turning_duration	0.022177
mean_turning_ang_acc	0.124055
std_turning_ang_acc	0.348202
pc25_turning_ang_acc	0.394237
pc75_turning_ang_acc	0.328811
skew_turning_ang_acc	0.279752
mean_turning_mag	0.399691
std_turning_mag	0.022188
pc25_turning_mag	0.522563
pc75_turning_mag	0.392625
skew_turning_mag	0.200364
mean_AI	0.728572
std_AI	0.396873
pc25_AI	0.620209
pc75_AI	0.744215
skew_AI	0.744248
mean_M1	0.730786
mean_M2	0.731342

**Tabella 4.1:** Coefficienti di Correlazione di Pearson tra features e diagnosi di Parkinson

### 4.1.2 Ottimizzazione degli Iperparametri e Classificazione

Affinché il confronto tra i quattro classificatori potesse essere il più possibile attendibile, si è scelto di effettuare un'ottimizzazione degli iperparametri, in modo da poter ottenere così la massima resa da ogni classificatore ed ottenere il valore migliore di F1-Score. Come già anticipato nel capitolo precedente, i classificatori scelti sono stati

- Support Vector Classifier
- K-Nearest Neighbour Classifier
- Random Forest Classifier
- Gradient Boosting Classifier

Per ogni classificatore si è scelto un set di valori attribuibili agli iperparametri principali, successivamente è stato creato un dizionario riportato nella tabella 4.1.2

```

1 model_params = {
2     'svm': {
3         'model': svm.SVC(gamma='auto'),
4         'params': {
5             'C': [1, 10, 20],
6             'kernel': ['rbf', 'linear']
7         }
8     },
9     'KNN': {
10        'model': KNeighborsClassifier(),
11        'params': {
12            'n_neighbors': [1, 10, 2],
13            'weights': ['uniform', 'distance'],
14            'metric': ['euclidean', 'manhattan', 'minkowski'],
15            'leaf_size': range(1, 50, 5)
16        }
17    },
18    'RF': {
19        'model': RandomForestClassifier(),
20        'params': {
21            'n_estimators': [10, 17, 25],
22            'max_features': [1, 3, 6],
23            'max_depth': [2, 4],
24            'min_samples_split': [2, 5],

```



```

25         'min_samples_leaf': [1, 2],
26         'bootstrap': [True, False]
27     },
28 },
29 'GB': {
30     'model': GradientBoostingClassifier(),
31     'params': {
32         'n_estimators': [10, 17, 25],
33         'learning_rate': [0.001, 0.01, 0.1],
34         'max_depth': [2, 4],
35         'max_features': [1, 3, 6],
36     }
37 },
38 }

```

**Tabella 4.2:** Dizionario per ottimizzazione Iperparametri

Il dizionario creato è poi stato utilizzato per l'ottimizzazione basata su F1-Score che ha dato come risultato i valori indicati nella tabella 4.2, a loro volta utilizzati per configurare i classificatori.

```

1 SVC = {'C': 1, 'kernel': 'linear'}
2
3 KNN = {'leaf_size': 1, 'metric': 'euclidean', 'n_neighbors': 1, '
      weights': 'uniform'}
4
5 RF = {'bootstrap': True, 'max_depth': 2, 'max_features': 3, '
      min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 17}
6
7 GB = {'learning_rate': 0.1, 'max_depth': 2, 'max_features': 1, '
      n_estimators': 25}

```

**Tabella 4.3:** Risultato dell'ottimizzazione degli Iperparametri

Nella tabella 4.4 sono riportati i risultati ottenuti dalla classificazione, i quali mettono in risalto la prevalenza del classificatore basato su Support Vector Machines con una prevalenza sugli altri classificatori prendendo come riferimento il valore dell'F1-Score.

Modello	F1-Score	Accuratezza	Recall
SVC	0.91	0.95	0.83
KNN	0.80	0.90	0.67
RF	0.50	0.80	0.33
GB	0.67	0.85	0.50

**Tabella 4.4:** Risultati Classificazione

### 4.1.3 Analisi Statistica

Dopo aver esportato le features calcolate di tutti i pazienti, si è proseguito ad eseguire un'analisi statistica per determinare la validità statistica dei dati raccolti. Nella tabella ?? sono riportate tutte le Features, per ognuna due righe corrispondenti alla classe PD con valore 1 per i casi Parkinson e 0 per i campioni Sani. Le colonne della tabella indicano, in ordine da sinistra verso destra, rispettivamente:

- Nome feature
- Classe di appartenenza PD
- Numero di campioni della classe
- Campioni mancanti
- Media dei valori
- Mediana dei valori
- Deviazione standard dei valori
- Valore minimo
- Valore massimo

Complessivamente, i valori ottenuti si sono rivelati statisticamente significativi, motivo per cui si è proseguito ad effettuare dei test statistici per valutarne la rilevanza.

## Risultati

Descrittive		GOOD_SLEEPER	N	Mancanti	Media	Mediana	SD	Minimo	Massimo
tot_hrs	0		19	0	7.8064	8.03010	1.25610	5.13156	10.3802
	1		13	0	7.0907	7.25318	1.23138	4.58193	8.7643
mean_tilt_angle	0		19	0	65.4617	64.96740	10.30799	49.35398	84.1882
	1		13	0	68.6641	68.58819	8.45101	52.51156	83.2976
std_tilt_angle	0		19	0	21.1376	21.39044	7.27524	9.76798	35.7358
	1		13	0	21.8147	21.62819	7.07996	6.49448	35.0103
pc25_tilt_angle	0		19	0	51.4671	49.81258	18.45938	13.99045	83.3093
	1		13	0	55.3447	53.90498	16.26744	19.48155	80.8216
pc75_tilt_angle	0		19	0	82.1104	85.42364	5.94101	68.01315	88.9109
	1		13	0	85.2754	85.52861	2.85679	79.86662	88.7024
skew_tilt_angle	0		19	0	-1.3031	-0.99904	1.24994	-5.06566	-0.1916
	1		13	0	-1.5147	-1.09514	1.34656	-5.45864	-0.0405
number_turns	0		19	0	14.8947	14	7.71647	4	29
	1		13	0	13.8462	12	5.85728	7	25
mean_turning_interval	0		19	0	32.9539	30.83333	12.96796	16.83929	62.7000
	1		13	0	32.2094	28.68182	14.01984	19.38095	75.0000
std_turning_interval	0		19	0	27.5780	22.62899	12.53538	10.26921	60.7731
	1		13	0	25.2890	22.27564	10.05972	14.60873	51.9519
pc25_turning_interval	0		19	0	14.5066	12.50000	11.22540	4.75000	52.0000
	1		13	0	14.3654	10.00000	11.70764	6.00000	50.8750
pc75_turning_interval	0		19	0	42.2500	38.37500	19.39748	19.50000	85.0000
	1		13	0	44.1442	46.00000	13.54466	23.37500	77.0000
skew_turning_interval	0		19	0	1.2659	1.11886	0.78206	-0.10926	2.6838
	1		13	0	0.9035	1.03158	0.49840	0.02967	1.5662
mean_turning_ang_vel	0		19	0	61.8246	58.73564	24.85646	25.48358	117.9072
	1		13	0	66.3132	66.09896	21.14329	33.36287	100.9619
std_turning_ang_vel	0		19	0	47.3860	38.97203	28.52061	12.70320	134.4294
	1		13	0	40.0233	36.44082	20.36851	8.10005	73.3490
pc25_turning_ang_vel	0		19	0	29.2542	23.00928	19.79753	10.10653	77.9671
	1		13	0	38.0444	38.78367	14.83660	16.61780	65.5903
pc75_turning_ang_vel	0		19	0	75.3283	84.28347	36.19357	17.99287	142.8781
	1		13	0	91.9436	86.37432	36.89734	38.06295	164.3060
skew_turning_ang_vel	0		19	0	0.9412	0.89773	1.06227	-1.01144	2.9974
	1		13	0	0.5904	0.55666	0.87496	-0.43532	3.1798
mean_turning_duration	0		19	0	0.4321	0.42299	0.12631	0.19139	0.7237
	1		13	0	0.5312	0.50648	0.12961	0.38242	0.8583
std_turning_duration	0		19	0	0.2407	0.21141	0.11009	0.09553	0.5292
	1		13	0	0.3026	0.26076	0.09362	0.16773	0.5023
pc25_turning_duration	0		19	0	0.2634	0.24838	0.09757	0.09617	0.3993
	1		13	0	0.3343	0.29563	0.14086	0.18022	0.7011
pc75_turning_duration	0		19	0	0.5363	0.49385	0.16191	0.22679	0.8898
	1		13	0	0.6681	0.71306	0.16641	0.40939	1.0372
skew_turning_duration	0		19	0	0.6364	0.59262	0.63702	-0.40832	1.6633
	1		13	0	1.1046	1.21589	0.62363	-0.38692	1.9546
mean_turning_ang_acc	0		19	0	265.8610	230.94787	172.96605	84.16734	642.7275
	1		13	0	206.3730	256.02723	100.49090	43.20775	326.7300
std_turning_ang_acc	0		19	0	357.3851	203.04471	407.18127	54.10413	1410.9029
	1		13	0	202.6489	173.66051	184.03482	31.89878	729.6488
pc25_turning_ang_acc	0		19	0	71.4482	55.12135	55.15106	18.06810	186.7588

Descrittive		GOOD_SLEEPER	N	Mancanti	Media	Mediana	SD	Minimo	Massimo
	1		13	0	68.5006	63.03726	37.05016	21.03579	143.3144
pc75_turning_ang_acc	0		19	0	253.3014	249.53916	118.58088	39.37663	452.8138
	1		13	0	275.5515	283.50830	152.04837	52.90472	511.4005
skew_turning_ang_acc	0		19	0	1.5657	1.35801	1.14247	-0.10517	4.1847
	1		13	0	1.0627	0.84516	1.07584	-0.42341	4.1425
mean_turning_mag	0		19	0	0.0301	0.02847	0.01440	0.00413	0.0570
	1		13	0	0.0381	0.03451	0.01030	0.01882	0.0531
std_turning_mag	0		19	0	0.0277	0.02828	0.01502	0.00357	0.0550
	1		13	0	0.0317	0.03262	0.01216	0.01213	0.0509
pc25_turning_mag	0		19	0	0.0102	0.00746	0.00860	9.60e-4	0.0305
	1		13	0	0.0138	0.01253	0.00847	0.00261	0.0350
pc75_turning_mag	0		19	0	0.0426	0.04422	0.02055	0.00484	0.0775
	1		13	0	0.0519	0.05026	0.01624	0.02876	0.0928
skew_turning_mag	0		19	0	0.8096	0.68585	0.65756	-0.43716	1.9715
	1		13	0	0.9623	1.00458	0.61649	-0.03925	1.7899
mean_AI	0		19	0	0.2222	0.13033	0.17944	0.09317	0.5939
	1		13	0	0.1316	0.11753	0.03777	0.09245	0.2268
std_AI	0		19	0	0.6301	0.57282	0.19349	0.41813	1.0530
	1		13	0	0.5833	0.55202	0.12676	0.42639	0.8616
pc25_AI	0		19	0	0.0335	0.02019	0.03022	0.01448	0.1327
	1		13	0	0.0209	0.02124	0.00210	0.01675	0.0245
pc75_AI	0		19	0	0.1472	0.02918	0.20683	0.02260	0.5723
	1		13	0	0.0278	0.02539	0.00670	0.02366	0.0485
skew_AI	0		19	0	6.6856	7.27784	2.06005	3.08913	8.9254
	1		13	0	7.6159	7.58097	1.28617	5.48721	9.8232
mean_M1	0		19	0	0.2194	0.12768	0.17922	0.08846	0.5889
	1		13	0	0.1279	0.11426	0.03726	0.08764	0.2214
mean_M2	0		19	0	0.2176	0.12713	0.17847	0.08508	0.5839
	1		13	0	0.1257	0.11374	0.03686	0.08609	0.2181

Tabella 4.5: Statistiche descrittive Sani-PD

Il primo test che si è effettuato è stato il test di normalità Shapiro-Wilk, in cui vengono verificate le ipotesi statistiche. Nella tabella 4.6 sono state evidenziate in giallo le features in cui il criterio di normalità non viene rispettato. Per le variabili in cui l'ipotesi di normalità fosse stata confermata, si è proseguito con test T di Student a campioni indipendenti, che permette di determinare se vi è una differenza statisticamente significativa tra le medie di due gruppi tra loro indipendenti. Nella tabella 4.7 sono stati evidenziati in verde le features che non rispettano i criteri del test T di Student. Il test di Levene risulta essere significativo nei casi in cui il valore  $p < 0.05$ , avendo di conseguenza una violazione dell'assunzione di varianze uguali. Per le features in cui si è riscontrata tale violazione, si è effettuato il test U di Mann-Whitney. Nella tabella 4.8 sono riportate in giallo le features in cui il test T di Student non è superato, e in verde le variabili che non risultano idonee neppure al test U di Mann-Whitney.

Test di Normalità (Shapiro-Wilk)

	W	p
tot_hrs	0.990	0.990
mean_tilt_angle	0.969	0.465
std_tilt_angle	0.988	0.967
pc25_tilt_angle	0.964	0.344
pc75_tilt_angle	0.944	0.095
skew_tilt_angle	0.767	<.001
number_turns	0.958	0.240
mean_turning_interval	0.867	<.001
std_turning_interval	0.880	0.002
pc25_turning_interval	0.710	<.001
pc75_turning_interval	0.937	0.061
skew_turning_interval	0.964	0.348
mean_turning_ang_vel	0.971	0.522
std_turning_ang_vel	0.904	0.008
pc25_turning_ang_vel	0.968	0.458
pc75_turning_ang_vel	0.963	0.323
skew_turning_ang_vel	0.904	0.008
mean_turning_duration	0.959	0.261
std_turning_duration	0.952	0.160
pc25_turning_duration	0.869	0.001
pc75_turning_duration	0.961	0.288
skew_turning_duration	0.960	0.272
mean_turning_ang_acc	0.920	0.020
std_turning_ang_acc	0.855	<.001
pc25_turning_ang_acc	0.880	0.002
pc75_turning_ang_acc	0.927	0.033
skew_turning_ang_acc	0.876	0.002
mean_turning_mag	0.985	0.921
std_turning_mag	0.977	0.695
pc25_turning_mag	0.935	0.055
pc75_turning_mag	0.937	0.061
skew_turning_mag	0.947	0.117
mean_AI	0.888	0.003
std_AI	0.951	0.153
pc25_AI	0.682	<.001
pc75_AI	0.728	<.001
skew_AI	0.962	0.312
mean_M1	0.887	0.003
mean_M2	0.887	0.003

*Nota.* Un piccolo valore di p suggerisce una violazione del presupposto di normalità

**Tabella 4.6:** Test di Normalità di Shapiro-Wilk su Sani-PD

Test t a campioni indipendenti

		Statistiche	gdl	p
tot_hrs	t di Student	-1.381	30.0	0.177
mean_tilt_angle	t di Student	2.441	30.0	0.021
std_tilt_angle	t di Student	-0.603 <sup>a</sup>	30.0	0.551
pc25_tilt_angle	t di Student	1.065	30.0	0.295
pc75_tilt_angle	t di Student	5.683 <sup>a</sup>	30.0	<.001
skew_tilt_angle	t di Student	-1.013	30.0	0.319
number_turns	t di Student	-0.656 <sup>a</sup>	30.0	0.517
mean_turning_interval	t di Student	0.837	30.0	0.409
std_turning_interval	t di Student	-0.834 <sup>a</sup>	30.0	0.411
pc25_turning_interval	t di Student	1.590	30.0	0.122
pc75_turning_interval	t di Student	0.753	30.0	0.457
skew_turning_interval	t di Student	-0.842	30.0	0.407
mean_turning_ang_vel	t di Student	3.798	30.0	<.001
std_turning_ang_vel	t di Student	-0.413 <sup>a</sup>	30.0	0.683
pc25_turning_ang_vel	t di Student	3.586 <sup>a</sup>	30.0	0.001
pc75_turning_ang_vel	t di Student	5.676 <sup>a</sup>	30.0	<.001
skew_turning_ang_vel	t di Student	-2.580	30.0	0.015
mean_turning_duration	t di Student	0.451 <sup>a</sup>	30.0	0.655
std_turning_duration	t di Student	-0.416	30.0	0.680
pc25_turning_duration	t di Student	1.028 <sup>a</sup>	30.0	0.312
pc75_turning_duration	t di Student	0.310 <sup>a</sup>	30.0	0.759
skew_turning_duration	t di Student	-0.121	30.0	0.904
mean_turning_ang_acc	t di Student	-0.685 <sup>a</sup>	30.0	0.499
std_turning_ang_acc	t di Student	-2.035 <sup>a</sup>	30.0	0.051
pc25_turning_ang_acc	t di Student	2.350 <sup>a</sup>	30.0	0.026
pc75_turning_ang_acc	t di Student	1.907	30.0	0.066
skew_turning_ang_acc	t di Student	-1.596	30.0	0.121
mean_turning_mag	t di Student	2.388	30.0	0.023
std_turning_mag	t di Student	0.122 <sup>a</sup>	30.0	0.904
pc25_turning_mag	t di Student	3.357 <sup>a</sup>	30.0	0.002
pc75_turning_mag	t di Student	2.338	30.0	0.026
skew_turning_mag	t di Student	-1.120	30.0	0.272
mean_AI	t di Student	-5.826 <sup>a</sup>	30.0	<.001
std_AI	t di Student	-2.368 <sup>a</sup>	30.0	0.025
pc25_AI	t di Student	-4.331 <sup>a</sup>	30.0	<.001
pc75_AI	t di Student	-6.103 <sup>a</sup>	30.0	<.001
skew_AI	t di Student	6.103 <sup>a</sup>	30.0	<.001
mean_M1	t di Student	-5.864 <sup>a</sup>	30.0	<.001
mean_M2	t di Student	-5.873 <sup>a</sup>	30.0	<.001

Nota.  $H_a: \mu_0 \neq \mu_1$ <sup>a</sup> Il test di Levene è significativo ( $p < .05$ ), suggerendo una violazione dell'assunzione di varianze uguali

Tabella 4.7: Test T di Student su Sani-PD

Test t a campioni indipendenti

		Statistiche	p
tot_hrs	U di Mann-Whitney	60.0	0.122
mean_tilt_angle	U di Mann-Whitney	42.0	0.020
std_tilt_angle	U di Mann-Whitney	83.0	0.586
pc25_tilt_angle	U di Mann-Whitney	63.0	0.157
pc75_tilt_angle	U di Mann-Whitney	14.0	<.001
skew_tilt_angle	U di Mann-Whitney	88.0	0.744
number_turns	U di Mann-Whitney	89.5	0.794
mean_turning_interval	U di Mann-Whitney	70.0	0.267
std_turning_interval	U di Mann-Whitney	92.0	0.879
pc25_turning_interval	U di Mann-Whitney	47.5	0.037
pc75_turning_interval	U di Mann-Whitney	77.0	0.421
skew_turning_interval	U di Mann-Whitney	82.0	0.557
mean_turning_ang_vel	U di Mann-Whitney	23.0	0.002
std_turning_ang_vel	U di Mann-Whitney	77.0	0.421
pc25_turning_ang_vel	U di Mann-Whitney	17.0	<.001
pc75_turning_ang_vel	U di Mann-Whitney	0.0	<.001
skew_turning_ang_vel	U di Mann-Whitney	36.0	0.010
mean_turning_duration	U di Mann-Whitney	82.0	0.557
std_turning_duration	U di Mann-Whitney	82.0	0.557
pc25_turning_duration	U di Mann-Whitney	54.0	0.071
pc75_turning_duration	U di Mann-Whitney	81.0	0.528
skew_turning_duration	U di Mann-Whitney	91.0	0.845
mean_turning_ang_acc	U di Mann-Whitney	87.0	0.711
std_turning_ang_acc	U di Mann-Whitney	83.0	0.586
pc25_turning_ang_acc	U di Mann-Whitney	37.0	0.011
pc75_turning_ang_acc	U di Mann-Whitney	53.0	0.064
skew_turning_ang_acc	U di Mann-Whitney	56.0	0.086
mean_turning_mag	U di Mann-Whitney	46.0	0.031
std_turning_mag	U di Mann-Whitney	94.0	0.948
pc25_turning_mag	U di Mann-Whitney	18.0	<.001
pc75_turning_mag	U di Mann-Whitney	36.0	0.010
skew_turning_mag	U di Mann-Whitney	77.0	0.421
mean_AI	U di Mann-Whitney	19.0	<.001
std_AI	U di Mann-Whitney	68.0	0.231
pc25_AI	U di Mann-Whitney	32.0	0.006
pc75_AI	U di Mann-Whitney	15.0	<.001
skew_AI	U di Mann-Whitney	11.0	<.001
mean_M1	U di Mann-Whitney	18.0	<.001
mean_M2	U di Mann-Whitney	20.0	0.001

Nota.  $H_a: \mu_0 \neq \mu_1$ 

Tabella 4.8: Test U di Mann-Whitney su Sani-PD

## 4.2 Classificazione Good/Bad Sleepers

### 4.2.1 Feature Selection

Riprendendo le procedure scelte per la classificazione Sani vs PD, le features sono state selezionate basandosi sul Coefficiente di Correlazione di Pearson in relazione alla feature Good\_Sleeper, ottenuta dal calcolo dei risultati dei questionari sPSQI e SLEEPS. Le features che si è scelto di mantenere le features con un valore del Coefficiente di Correlazione di Pearson strettamente maggiore a 0.3, nella tabella 4.9 sono riportati tutti i valori di correlazione, tra questi sono stati evidenziati i valori che si trovano al di sotto della soglia stabilita che hanno comportato la rimozione della corrispondente feature dalla classificazione.



Feature	Correlazione
tot_hrs	0.279690
tot_hrs	0.279690
mean_tilt_angle	0.166693
std_tilt_angle	0.047659
pc25_tilt_angle	0.110967
pc75_tilt_angle	0.308850
skew_tilt_angle	0.082942
number_turns	0.075422
mean_turning_interval	0.028174
std_turning_interval	0.099523
pc25_turning_interval	0.006271
pc75_turning_interval	0.055468
skew_turning_interval	0.259869
mean_turning_ang_vel	0.096669
std_turning_ang_vel	0.144503
pc25_turning_ang_vel	0.240716
pc75_turning_ang_vel	0.225119
skew_turning_ang_vel	0.176637
mean_turning_duration	0.366440
std_turning_duration	0.289497
pc25_turning_duration	0.294341
pc75_turning_duration	0.377997
skew_turning_duration	0.351874
mean_turning_ang_acc	0.199399
std_turning_ang_acc	0.227349
pc25_turning_ang_acc	0.030672
pc75_turning_ang_acc	0.084565
skew_turning_ang_acc	0.222820
mean_turning_mag	0.300117
std_turning_mag	0.145864
pc25_turning_mag	0.205462
pc75_turning_mag	0.240839
skew_turning_mag	0.119833
mean_AI	0.309759
std_AI	0.138393
pc25_AI	0.262072
pc75_AI	0.353567
skew_AI	0.254759
mean_M1	0.313248
mean_M2	0.315522

**Tabella 4.9:** Coefficienti di Correlazione di Pearson tra features e diagnosi di Parkinson

## 4.2.2 Ottimizzazione degli Iperparametri e Classificazione

La procedura di ottimizzazione degli Iperparametri è stata la medesima utilizzata per la Classificazione Sani PD. Anche in questo approccio l'ottimizzazione è stata impostata prendendo come riferimento il valore dell'F1-Score. Come già anticipato, i classificatori scelti sono stati

- Support Vector Classifier
- K-Nearest Neighbour Classifier
- Random Forest Classifier
- Gradient Boosting Classifier

Per l'ottimizzazione è stato riutilizzato il dizionario 4.1.2 utilizzato in precedenza. I risultati ottenuti sono riportati nella tabella 4.2.2, che successivamente sono stati a loro volta utilizzati per configurare i classificatori.

```

1 svm = { 'C': 1, 'kernel': 'linear' }
2
3 KNN = { 'leaf_size': 1, 'metric': 'euclidean', 'n_neighbors': 2, '
      weights': 'uniform' }
4
5 RF = { 'bootstrap': True, 'max_depth': 2, 'max_features': 6, '
      min_samples_leaf': 2, 'min_samples_split': 5, 'n_estimators': 17 }
6
7 GB = { 'learning_rate': 0.1, 'max_depth': 2, 'max_features': 6, '
      n_estimators': 17 }

```

**Tabella 4.10:** Risultato dell'ottimizzazione degli Iperparametri

Nella tabella 4.4 sono riportati i risultati ottenuti dalla classificazione, i quali mettono in risalto la prevalenza del classificatore basato su Support Vector Machines con una prevalenza sugli altri classificatori prendendo come riferimento il valore dell'F1-Score.

Modello	F1-Score	Accuratezza	Recall
SVC	0.95	0.95	1.00
KNN	0.43	0.60	0.33
RF	0.89	0.90	0.89
GB	0.82	0.85	0.78

**Tabella 4.11:** Risultati Classificazione Good Sleepers

### 4.2.3 Analisi Statistica

Concluso il processo di classificazione si è scelto anche in questo caso di proseguire con un'analisi statistica per determinare la validità statistica dei dati raccolti. Nella tabella 4.12 sono riportate tutte le Features, per ognuna due righe corrispondenti alla classe Good\_Sleeper con valore 1 per i soggetti dai quali tramite i questionari si è dedotta una buona qualità del sonno e 0 per i soggetti che hanno mostrato, sempre secondo i questionari, una bassa qualità del sonno. Le colonne della tabella indicano, in ordine da sinistra verso destra, rispettivamente:

- Nome feature
- Classe di appartenenza PD
- Numero di campioni della classe
- Campioni mancanti
- Media dei valori
- Mediana dei valori
- Deviazione standard dei valori
- Valore minimo
- Valore massimo

Complessivamente, anche in questo caso i valori ottenuti si sono rivelati statisticamente significativi, motivo per cui si è proseguito ad effettuare dei test statistici per valutarne la rilevanza.

## Risultati

Descrittive								
	GOOD_SLEEPER	N	Mancanti	Media	Mediana	SD	Minimo	Massimo
tot_hrs	0	19	0	7.8064	8.03010	1.25610	5.13156	10.3802
	1	13	0	7.0907	7.25318	1.23138	4.58193	8.7643
mean_tilt_angle	0	19	0	65.4617	64.96740	10.30799	49.35398	84.1882
	1	13	0	68.6641	68.58819	8.45101	52.51156	83.2976
std_tilt_angle	0	19	0	21.1376	21.39044	7.27524	9.76798	35.7358
	1	13	0	21.8147	21.62819	7.07996	6.49448	35.0103
pc25_tilt_angle	0	19	0	51.4671	49.81258	18.45938	13.99045	83.3093
	1	13	0	55.3447	53.90498	16.26744	19.48155	80.8216
pc75_tilt_angle	0	19	0	82.1104	85.42364	5.94101	68.01315	88.9109
	1	13	0	85.2754	85.52861	2.85679	79.86662	88.7024
skew_tilt_angle	0	19	0	-1.3031	-0.99904	1.24994	-5.06566	-0.1916
	1	13	0	-1.5147	-1.09514	1.34656	-5.45864	-0.0405
number_turns	0	19	0	14.8947	14	7.71647	4	29
	1	13	0	13.8462	12	5.85728	7	25
mean_turning_interval	0	19	0	32.9539	30.83333	12.96796	16.83929	62.7000
	1	13	0	32.2094	28.68182	14.01984	19.38095	75.0000
std_turning_interval	0	19	0	27.5780	22.62899	12.53538	10.26921	60.7731
	1	13	0	25.2890	22.27564	10.05972	14.60873	51.9519
pc25_turning_interval	0	19	0	14.5066	12.50000	11.22540	4.75000	52.0000
	1	13	0	14.3654	10.00000	11.70764	6.00000	50.8750
pc75_turning_interval	0	19	0	42.2500	38.37500	19.39748	19.50000	85.0000
	1	13	0	44.1442	46.00000	13.54466	23.37500	77.0000
skew_turning_interval	0	19	0	1.2659	1.11886	0.78206	-0.10926	2.6838
	1	13	0	0.9035	1.03158	0.49840	0.02967	1.5662
mean_turning_ang_vel	0	19	0	61.8246	58.73564	24.85646	25.48358	117.9072
	1	13	0	66.3132	66.09896	21.14329	33.36287	100.9619
std_turning_ang_vel	0	19	0	47.3860	38.97203	28.52061	12.70320	134.4294
	1	13	0	40.0233	36.44082	20.36851	8.10005	73.3490
pc25_turning_ang_vel	0	19	0	29.2542	23.00928	19.79753	10.10653	77.9671
	1	13	0	38.0444	38.78367	14.83660	16.61780	65.5903
pc75_turning_ang_vel	0	19	0	75.3283	84.28347	36.19357	17.99287	142.8781
	1	13	0	91.9436	86.37432	36.89734	38.06295	164.3060
skew_turning_ang_vel	0	19	0	0.9412	0.89773	1.06227	-1.01144	2.9974
	1	13	0	0.5904	0.55666	0.87496	-0.43532	3.1798
mean_turning_duration	0	19	0	0.4321	0.42299	0.12631	0.19139	0.7237
	1	13	0	0.5312	0.50648	0.12961	0.38242	0.8583
std_turning_duration	0	19	0	0.2407	0.21141	0.11009	0.09553	0.5292
	1	13	0	0.3026	0.26076	0.09362	0.16773	0.5023
pc25_turning_duration	0	19	0	0.2634	0.24838	0.09757	0.09617	0.3993
	1	13	0	0.3343	0.29563	0.14086	0.18022	0.7011
pc75_turning_duration	0	19	0	0.5363	0.49385	0.16191	0.22679	0.8898
	1	13	0	0.6681	0.71306	0.16641	0.40939	1.0372
skew_turning_duration	0	19	0	0.6364	0.59262	0.63702	-0.40832	1.6633
	1	13	0	1.1046	1.21589	0.62363	-0.38692	1.9546
mean_turning_ang_acc	0	19	0	265.8610	230.94787	172.96605	84.16734	642.7275
	1	13	0	206.3730	256.02723	100.49090	43.20775	326.7300
std_turning_ang_acc	0	19	0	357.3851	203.04471	407.18127	54.10413	1410.9029
	1	13	0	202.6489	173.66051	184.03482	31.89878	729.6488
pc25_turning_ang_acc	0	19	0	71.4482	55.12135	55.15106	18.06810	186.7588

Descrittive								
	GOOD_SLEEPER	N	Mancanti	Media	Mediana	SD	Minimo	Massimo
	1	13	0	68.5006	63.03726	37.05016	21.03579	143.3144
pc75_turning_ang_acc	0	19	0	253.3014	249.53916	118.58088	39.37663	452.8138
	1	13	0	275.5515	283.50830	152.04837	52.90472	511.4005
skew_turning_ang_acc	0	19	0	1.5657	1.35801	1.14247	-0.10517	4.1847
	1	13	0	1.0627	0.84516	1.07584	-0.42341	4.1425
mean_turning_mag	0	19	0	0.0301	0.02847	0.01440	0.00413	0.0570
	1	13	0	0.0381	0.03451	0.01030	0.01882	0.0531
std_turning_mag	0	19	0	0.0277	0.02828	0.01502	0.00357	0.0550
	1	13	0	0.0317	0.03262	0.01216	0.01213	0.0509
pc25_turning_mag	0	19	0	0.0102	0.00746	0.00860	9.60e-4	0.0305
	1	13	0	0.0138	0.01253	0.00847	0.00261	0.0350
pc75_turning_mag	0	19	0	0.0426	0.04422	0.02055	0.00484	0.0775
	1	13	0	0.0519	0.05026	0.01624	0.02876	0.0928
skew_turning_mag	0	19	0	0.8096	0.68585	0.65756	-0.43716	1.9715
	1	13	0	0.9623	1.00458	0.61649	-0.03925	1.7899
mean_AI	0	19	0	0.2222	0.13033	0.17944	0.09317	0.5939
	1	13	0	0.1316	0.11753	0.03777	0.09245	0.2268
std_AI	0	19	0	0.6301	0.57282	0.19349	0.41813	1.0530
	1	13	0	0.5833	0.55202	0.12676	0.42639	0.8616
pc25_AI	0	19	0	0.0335	0.02019	0.03022	0.01448	0.1327
	1	13	0	0.0209	0.02124	0.00210	0.01675	0.0245
pc75_AI	0	19	0	0.1472	0.02918	0.20683	0.02260	0.5723
	1	13	0	0.0278	0.02539	0.00670	0.02366	0.0485
skew_AI	0	19	0	6.6856	7.27784	2.06005	3.08913	8.9254
	1	13	0	7.6159	7.58097	1.28617	5.48721	9.8232
mean_M1	0	19	0	0.2194	0.12768	0.17922	0.08846	0.5889
	1	13	0	0.1279	0.11426	0.03726	0.08764	0.2214
mean_M2	0	19	0	0.2176	0.12713	0.17847	0.08508	0.5839
	1	13	0	0.1257	0.11374	0.03686	0.08609	0.2181

Tabella 4.12: Statistiche descrittive Good/Bad Sleepers

L'ordine ripetuto è stato lo stesso precedentemente applicato, in sequenza è stata verificata prima la normalità attraverso il test di Shapiro-Wilk, evidenziando in giallo nella tabella 4.13 le features che non rispettano il criterio, in seguito le features sono state sottoposte al test t di student come illustrato nella tabella 4.14, in cui sono evidenziate in verde le statistiche che non soddisfano i criteri del test, e, in seguito, le feature in questione sono state sottoposte al test U di Mann-Whitney, come riportato nella tabella 4.15, in cui sono evidenziate in verde le features che non superano neppure quest'ultimo test.

Test di Normalità (Shapiro-Wilk)

	W	p
tot_hrs	0.990	0.990
mean_tilt_angle	0.969	0.465
std_tilt_angle	0.988	0.967
pc25_tilt_angle	0.964	0.344
pc75_tilt_angle	0.944	0.095
skew_tilt_angle	0.767	<.001
number_turns	0.958	0.240
mean_turning_interval	0.867	<.001
std_turning_interval	0.880	0.002
pc25_turning_interval	0.710	<.001
pc75_turning_interval	0.937	0.061
skew_turning_interval	0.964	0.348
mean_turning_ang_vel	0.971	0.522
std_turning_ang_vel	0.904	0.008
pc25_turning_ang_vel	0.968	0.458
pc75_turning_ang_vel	0.963	0.323
skew_turning_ang_vel	0.904	0.008
mean_turning_duration	0.959	0.261
std_turning_duration	0.952	0.160
pc25_turning_duration	0.869	0.001
pc75_turning_duration	0.961	0.288
skew_turning_duration	0.960	0.272
mean_turning_ang_acc	0.920	0.020
std_turning_ang_acc	0.855	<.001
pc25_turning_ang_acc	0.880	0.002
pc75_turning_ang_acc	0.927	0.033
skew_turning_ang_acc	0.876	0.002
mean_turning_mag	0.985	0.921
std_turning_mag	0.977	0.695
pc25_turning_mag	0.935	0.055
pc75_turning_mag	0.937	0.061
skew_turning_mag	0.947	0.117
mean_AI	0.888	0.003
std_AI	0.951	0.153
pc25_AI	0.682	<.001
pc75_AI	0.728	<.001
skew_AI	0.962	0.312
mean_M1	0.887	0.003
mean_M2	0.887	0.003

*Nota.* Un piccolo valore di p suggerisce una violazione del presupposto di normalità

**Tabella 4.13:** Test di Normalità di Shapiro-Wilk su Good/Bad Sleepers

Test t a campioni indipendenti

		Statistiche	gdl	p
tot_hrs	t di Student	1.5956	30.0	0.121
mean_tilt_angle	t di Student	-0.9260	30.0	0.362
std_tilt_angle	t di Student	-0.2613	30.0	0.796
pc25_tilt_angle	t di Student	-0.6116	30.0	0.545
pc75_tilt_angle	t di Student	-1.7786 <sup>a</sup>	30.0	0.085
skew_tilt_angle	t di Student	0.4559	30.0	0.652
number_turns	t di Student	0.4143	30.0	0.682
mean_turning_interval	t di Student	0.1544	30.0	0.878
std_turning_interval	t di Student	0.5478	30.0	0.588
pc25_turning_interval	t di Student	0.0343	30.0	0.973
pc75_turning_interval	t di Student	-0.3043	30.0	0.763
skew_turning_interval	t di Student	1.4740	30.0	0.151
mean_turning_ang_vel	t di Student	-0.5320	30.0	0.599
std_turning_ang_vel	t di Student	0.7999	30.0	0.430
pc25_turning_ang_vel	t di Student	-1.3584	30.0	0.184
pc75_turning_ang_vel	t di Student	-1.2655	30.0	0.215
skew_turning_ang_vel	t di Student	0.9829	30.0	0.333
mean_turning_duration	t di Student	-2.1571	30.0	0.039
std_turning_duration	t di Student	-1.6566	30.0	0.108
pc25_turning_duration	t di Student	-1.6869	30.0	0.102
pc75_turning_duration	t di Student	-2.2363	30.0	0.033
skew_turning_duration	t di Student	-2.0590	30.0	0.048
mean_turning_ang_acc	t di Student	1.1145	30.0	0.274
std_turning_ang_acc	t di Student	1.2787	30.0	0.211
pc25_turning_ang_acc	t di Student	0.1681	30.0	0.868
pc75_turning_ang_acc	t di Student	-0.4648	30.0	0.645
skew_turning_ang_acc	t di Student	1.2519	30.0	0.220
mean_turning_mag	t di Student	-1.7232	30.0	0.095
std_turning_mag	t di Student	-0.8076	30.0	0.426
pc25_turning_mag	t di Student	-1.1499	30.0	0.259
pc75_turning_mag	t di Student	-1.3591	30.0	0.184
skew_turning_mag	t di Student	-0.6611	30.0	0.514
mean_AI	t di Student	1.7844 <sup>a</sup>	30.0	0.084
std_AI	t di Student	0.7654	30.0	0.450
pc25_AI	t di Student	1.4874 <sup>a</sup>	30.0	0.147
pc75_AI	t di Student	2.0703 <sup>a</sup>	30.0	0.047
skew_AI	t di Student	-1.4430	30.0	0.159
mean_M1	t di Student	1.8067 <sup>a</sup>	30.0	0.081
mean_M2	t di Student	1.8212 <sup>a</sup>	30.0	0.079

Nota.  $H_a: \mu_0 \neq \mu_1$ <sup>a</sup> Il test di Levene è significativo ( $p < .05$ ), suggerendo una violazione dell'assunzione di varianze uguali

Tabella 4.14: Test T di Student su Good/Bad Sleepers

Test t a campioni indipendenti

		Statistiche	p
tot_hrs	U di Mann-Whitney	85.0	0.145
mean_tilt_angle	U di Mann-Whitney	101.0	0.399
std_tilt_angle	U di Mann-Whitney	112.0	0.673
pc25_tilt_angle	U di Mann-Whitney	104.0	0.466
pc75_tilt_angle	U di Mann-Whitney	90.0	0.205
skew_tilt_angle	U di Mann-Whitney	102.0	0.420
number_turns	U di Mann-Whitney	115.5	0.773
mean_turning_interval	U di Mann-Whitney	118.0	0.848
std_turning_interval	U di Mann-Whitney	111.0	0.645
pc25_turning_interval	U di Mann-Whitney	120.5	0.924
pc75_turning_interval	U di Mann-Whitney	110.0	0.618
skew_turning_interval	U di Mann-Whitney	102.0	0.420
mean_turning_ang_vel	U di Mann-Whitney	109.0	0.591
std_turning_ang_vel	U di Mann-Whitney	113.0	0.701
pc25_turning_ang_vel	U di Mann-Whitney	79.0	0.091
pc75_turning_ang_vel	U di Mann-Whitney	95.0	0.283
skew_turning_ang_vel	U di Mann-Whitney	92.0	0.234
mean_turning_duration	U di Mann-Whitney	72.0	0.050
std_turning_duration	U di Mann-Whitney	72.0	0.050
pc25_turning_duration	U di Mann-Whitney	97.0	0.318
pc75_turning_duration	U di Mann-Whitney	70.0	0.042
skew_turning_duration	U di Mann-Whitney	70.0	0.042
mean_turning_ang_acc	U di Mann-Whitney	113.0	0.701
std_turning_ang_acc	U di Mann-Whitney	95.0	0.283
pc25_turning_ang_acc	U di Mann-Whitney	113.0	0.701
pc75_turning_ang_acc	U di Mann-Whitney	107.0	0.539
skew_turning_ang_acc	U di Mann-Whitney	82.0	0.116
mean_turning_mag	U di Mann-Whitney	75.0	0.065
std_turning_mag	U di Mann-Whitney	105.0	0.490
pc25_turning_mag	U di Mann-Whitney	84.0	0.135
pc75_turning_mag	U di Mann-Whitney	92.0	0.234
skew_turning_mag	U di Mann-Whitney	107.0	0.539
mean_AI	U di Mann-Whitney	94.0	0.266
std_AI	U di Mann-Whitney	110.0	0.618
pc25_AI	U di Mann-Whitney	118.0	0.848
pc75_AI	U di Mann-Whitney	92.0	0.234
skew_AI	U di Mann-Whitney	103.0	0.443
mean_M1	U di Mann-Whitney	96.0	0.300
mean_M2	U di Mann-Whitney	96.0	0.300

Nota.  $H_a: \mu_0 \neq \mu_1$ 

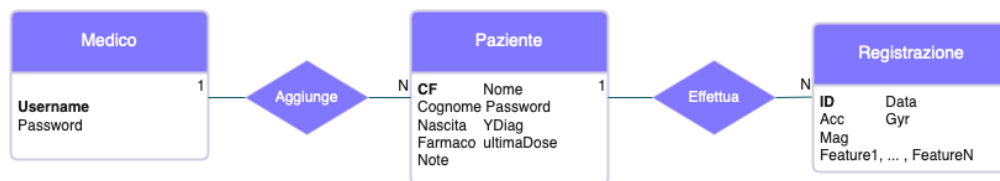
Tabella 4.15: Test U di Mann-Whitney su Good/Bad Sleepers



### 4.3 Verso la medicina pervasiva: Applicazione per Apple Watch<sup>®</sup>

Con l'innalzamento dell'età della popolazione, i servizi di home care sono sempre più richiesti e in alcuni casi necessari per abbattere problematiche legate alla raggiungibilità dei centri di cura. Con la rivoluzione dell'Industria 4.0, sta prendendo piede anche l'Healthcare 4.0, caratterizzato dall'home care costituito da soluzioni che combinano l'utilizzo dell'intelligenza artificiale con le soluzioni di Cloud Computing. [31]

La soluzione a cui si è pensato, a seguito del raggiungimento di buoni risultati con la classificazione basata su sensori inerziali presenti su dispositivi special purpose, è stata la progettazione di un'applicazione per dispositivi wearable general purpose (Smartwatch), in combinazione con una soluzione Cloud che possa permettere al medico curante di accedere ai risultati della classificazione o ottenere i dati dei sensori inerziali per estendere il processo di diagnostica. Nella figura 4.1 è rappresentato un diagramma Entità-Relazione rappresentativo del database che si è pensato di realizzare per l'implementazione dell'applicazione.



**Figura 4.1:** Diagramma Entità-Relazione per Database

#### 4.3.1 Applicazione Lato Paziente

L'applicazione lato paziente si basa su un'app mobile per Apple Watch il cui funzionamento è riassunto nella figura 4.2, in cui appunto il paziente, dopo aver fatto il login su smartphone ed essersi identificato, ha la possibilità di avviare la registrazione prima di andare a dormire direttamente dal proprio smartwatch, con la possibilità di interromperla in qualsiasi momento. Al termine della registrazione notturna il paziente può avere un primo feedback legato alla qualità del sonno grazie ad un punteggio, ottenuto grazie ad una somma pesata delle features registrate, queste ultime selezionate dalle rimanenti nella tabella 4.9, il cui peso è definito dal valore di correlazione normalizzato. Una volta conclusa la registrazione, oltre a mostrare il primo feedback al paziente, dal lato backend vengono salvate sul server i dati dei sensori inerziali e calcolate le features.



**Figura 4.2:** Prototipo di funzionamento app per Apple Watch

### 4.3.2 Applicazione Lato Medico

L'applicazione lato medico è una Web Application basata su un servizio Cloud di tipo SaaS (Software as a Service) in cui il medico accede con le proprie credenziali e successivamente, come mostrato nella figura 4.3, ha la possibilità di vedere tutti i propri pazienti aderenti allo studio ed eventualmente cercarli per nome tramite la barra di ricerca. Una volta selezionato il paziente di interesse, si apre una nuova sezione (illustrata nella figura 4.4) in cui oltre all'anagrafica del paziente e ad una breve sezione dedicata alle note del medico, si presenta nella sezione principale lo storico delle registrazioni tra le quali il medico può selezionarne una per avere il dettaglio. Una volta entrato nel dettaglio, il medico, come illustrato in figura 4.5, avrà una vista completa dei valori delle features attualmente implementate e i grafici relativi ad Accelerometro, Giroscopio e Magnetometro. Per ogni grafico si ha la possibilità di scaricare in formato .csv separatamente i valori dei sensori inerziali triassiali. Infine, vi è un riquadro che indica la previsione effettuata attraverso il classificatore, che indica se la qualità del sonno è cattiva e se si sospetta un caso di Parkinson.

trova paziente

vai

Cognome	Nome	Codice Fiscale	Data di nascita	
Rossi	Mario	RSSMRA54H12Y328A	12/05/1954	Apri Scheda
Camilo	Beatrice	CMLBTR44J10T738M	22/08/1996	Apri Scheda
Verdi	Giuseppe	VRDGSP13G10G837H	10/10/1813	Apri Scheda
Di Gangi	Luca	DGNLCU97K08R442I	08/06/1997	Apri Scheda
Sarais	Martina	SRSMT97M29T473I	29/05/1997	Apri Scheda
Chiotti	Giorgia	CHTGRG97J02M554I	02/12/1997	Apri Scheda
Fiallo	Amina	FLLMNA97K28B493A	28/10/1997	Apri Scheda
Orzylowska	Dominika	RZLDMN96B26D328I	26/11/1996	Apri Scheda
Barbieri	Bianca	BRBBNC97K07X528G	07/12/1997	Apri Scheda

Figura 4.3: Mockup Home

Scheda Paziente

Indietro

Anagrafica

Nome

Mario

Cognome

Rossi

Codice Fiscale

RSSMRA54H12Y328A

Nascita

12/05/1954

Anni diagnosi

5

Farmaco

Levodopa

Ultima dose

18:00

Note

13-04-2023

Apri dettagli

10-04-2023

Apri dettagli

09-04-2023

Apri dettagli

07-04-2023

Apri dettagli

05-04-2023

Apri dettagli

02-04-2023

Apri dettagli

Figura 4.4: Mockup Schermata paziente



Figura 4.5: Mockup schermata dettaglio

# Capitolo 5

## Discussione

### 5.1 Classificazione Sani PD

#### 5.1.1 Feature Selection

Grazie alla feature selection, i valori di Correlazione di Pearson hanno messo in risalto la presenza di una buona correlazione di diverse features. Le features che hanno mostrato maggiore correlazione tra tutte descrivono l'angolo di rotazione, la velocità angolare delle rotazioni e la variazione di campo magnetico durante le rotazioni, ciò sta a dimostrare la correlazione tra motilità notturna e diagnosi di parkinson.

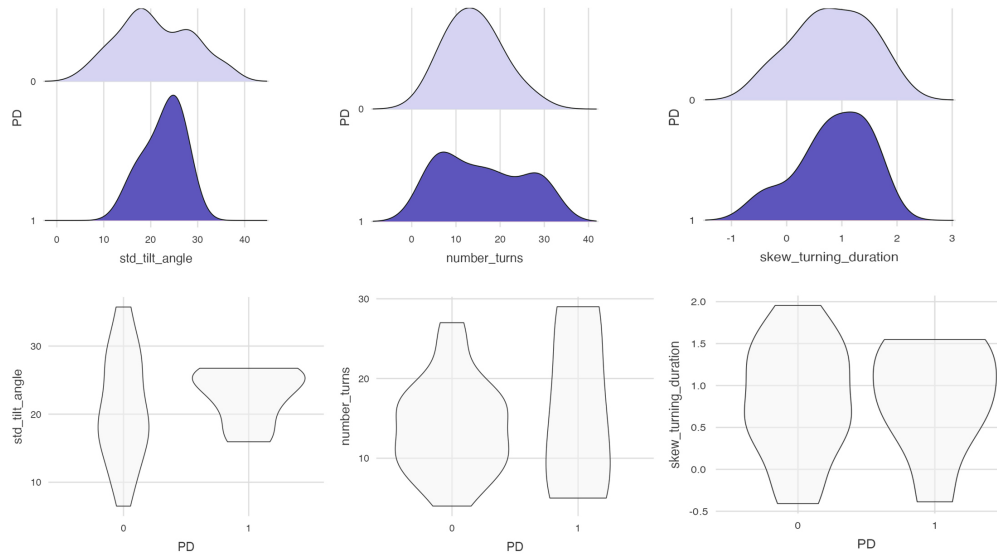
#### 5.1.2 Classificazione

I risultati ottenuti dalla classificazione hanno dimostrato l'efficienza della feature selection e la rilevanza delle features utilizzate. Il classificatore basato su Support Vector Machines è stato selezionato in quanto migliore tra i classificatori implementati, sia per il valore di F1-Score ottenuto, sia per i valori di Accuraztezza e Recall.

#### 5.1.3 Analisi Statistica

L'analisi effettuata ha dimostrato la rilevanza statistica delle features prodotte. Nello specifico, buona parte delle features sono distribuite normalmente e di conseguenza si sono potuti effettuare i test necessari a determinarne la validità statistica. Complessivamente anche il Test T di student e in seguito U di Mann-Whitney hanno riscontrato buoni risultati, dimostrando la rilevanza statistica del lavoro svolto. Sono state riportate in seguito nella figura 5.1 le distribuzioni e i

violin plot inerenti alle features su cui si sono riscontrati risultati migliori in termini di Feature selection e Analisi statistica.



**Figura 5.1:** Risultati statistiche delle features maggiormente rilevanti rispetto la classe Sani/PD

## 5.2 Classificazione Bad/Good Sleepers

### 5.2.1 Feature Selection

I valori ottenuti dalla correlazione tra le features e i questionari mostrano come vi sia correlazione principalmente tra la durata di rotazione e angolo di rotazione con la percezione della qualità del sonno.

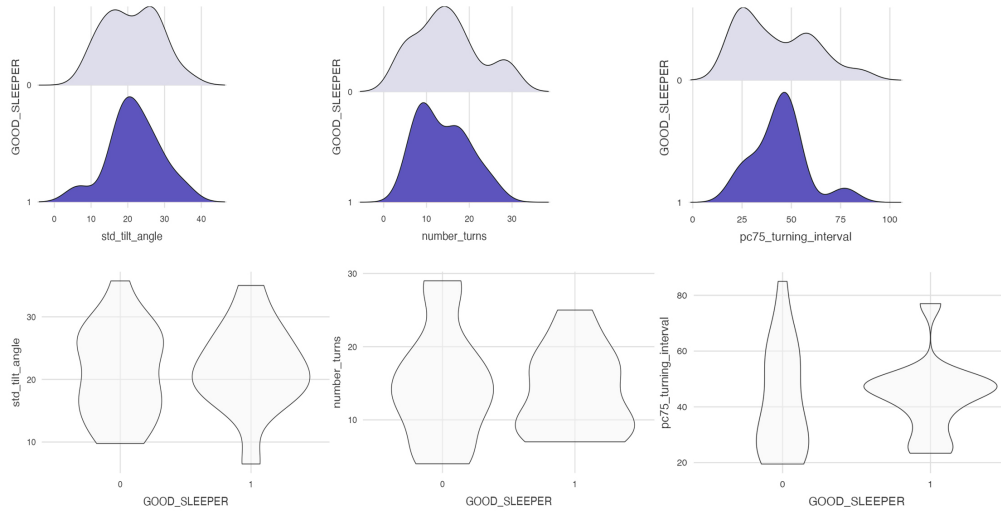
### 5.2.2 Classificazione

La classificazione anche in questo caso ha mostrato buoni risultati, il modello migliore anche in questo caso si è rivelato essere il classificatore basato su Support Vector Machines, con buoni risultati di Accuratezza e Recall.

### 5.2.3 Analisi Statistica

L'analisi in termini statistici ha dato buoni risultati anche nelle variabili per la classificazione Good/Bad Sleepers, mostrando di fatto una vicinanza tra disturbi

del sonno e percezione della qualità del sonno. Anche in questo caso buona parte delle features rispetta una distribuzione normale permettendo l'applicazione dei test per determinarne la validità statistica. La validità della rilevanza statistica è stata confermata anche in relazione alle classi Good/Bad Sleepers, dando ulteriore conferma di validità delle features realizzate. Nella figura 5.2 sono state riportate le statistiche delle features che hanno riscontrato risultati migliori in seguito ai test statistici e alla feature selection.



**Figura 5.2:** Risultati statistiche delle features maggiormente rilevanti rispetto la classe Bad/Good Sleepers

## Capitolo 6

# Conclusione e Sviluppi Futuri

il lavoro ha dimostrato buone potenzialità nella discriminazione di parkinson; nel futuro sarà opportuno guardare più nel dettaglio le tipologie di acinesia, quindi esplorare solo pazienti di PD, a diversi stadi della malattia.

In questo sarà utile sviluppare in maniera più strutturata una app per apple watch al fine di fornire ai pazienti una soluzione completamente pervasiva, indossabile e minimamente invasiva, come proposto nel capitolo 4, offrendo così, grazie all'utilizzo di una soluzione cloud, anche una maggiore accessibilità alle cure ed al monitoraggio per pazienti impossibilitati al raggiungimento dei centri di cura specializzati. Inoltre, la raccolta dati, come già anticipato, ha coinvolto anche la raccolta di campioni vocali, quindi nel futuro sarà esplorata una classificazione multimodale che coinvolgerà parametri motori per un'analisi inerziale e parametri vocali, seguendo l'approccio proposto nell'articolo scelto il questionario SLEEPS sulla qualità del sonno utilizzato in fase di raccolta dati[28].



# Bibliografia

- [1] Savitt JM Reich SG. «The Medical clinics of North America». In: 103(2) (dic. 2018), pp. 337–350 (cit. a p. 1).
- [2] Parkinson J. «An essay on the shaking palsy». In: (1817) (cit. a p. 1).
- [3] *Malattia idiopatica*. Nov. 2021. URL: [https://it.wikipedia.org/wiki/Malattia\\_idiopatica](https://it.wikipedia.org/wiki/Malattia_idiopatica) (cit. a p. 1).
- [4] Janice M Beitz. «Parkinson s disease a review». In: *Frontiers in Bioscience* S6.1 (2014), pp. 65–74. DOI: 10.2741/s415. URL: <https://doi.org/10.2741/s415> (cit. a p. 1).
- [5] *Statistics*. URL: <https://www.parkinson.org/understanding-parkinsons/statistics> (cit. a p. 1).
- [6] *Malattie del sistema extrapiramidale / ICS Maugeri — icsmaugeri.it*. <https://www.icsmaugeri.it/patologie/malattie-del-sistema-extrapiramidale>. [Accessed 06-Apr-2023] (cit. a p. 1).
- [7] G. DeMaagd e A. Philip. «Parkinson’s Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis». In: *P T* 40.8 (ago. 2015), pp. 504–532 (cit. a p. 1).
- [8] Lorraine V Kalia e Anthony E Lang. «Parkinson’s disease». In: *The Lancet* 386.9996 (ago. 2015), pp. 896–912. DOI: 10.1016/s0140-6736(14)61393-3. URL: [https://doi.org/10.1016/s0140-6736\(14\)61393-3](https://doi.org/10.1016/s0140-6736(14)61393-3) (cit. a p. 2).
- [9] Bastiaan R Bloem, Michael S Okun e Christine Klein. «Parkinson’s disease». In: *The Lancet* 397.10291 (giu. 2021), pp. 2284–2303. DOI: 10.1016/s0140-6736(21)00218-x. URL: [https://doi.org/10.1016/s0140-6736\(21\)00218-x](https://doi.org/10.1016/s0140-6736(21)00218-x) (cit. a p. 2).
- [10] Sigurlaug Sveinbjornsdottir. «The clinical symptoms of Parkinson’s disease». In: *Journal of Neurochemistry* 139 (lug. 2016), pp. 318–324. DOI: 10.1111/jnc.13691. URL: <https://doi.org/10.1111/jnc.13691> (cit. a p. 2).

- [11] J Jankovic. «Parkinson's disease: clinical features and diagnosis». In: *Journal of Neurology, Neurosurgery & Psychiatry* 79.4 (apr. 2008), pp. 368–376. DOI: 10.1136/jnnp.2007.131045. URL: <https://doi.org/10.1136/jnnp.2007.131045> (cit. a p. 2).
- [12] Peter A. LeWitt. «Levodopa therapy for Parkinson's disease: Pharmacokinetics and pharmacodynamics». In: *Movement Disorders* 30.1 (dic. 2014), pp. 64–72. DOI: 10.1002/mds.26082. URL: <https://doi.org/10.1002/mds.26082> (cit. a p. 3).
- [13] Carl E Clarke et al. «Clinical effectiveness and cost-effectiveness of physiotherapy and occupational therapy versus no therapy in mild to moderate Parkinson's disease: a large pragmatic randomised controlled trial (PD REHAB)». In: *Health Technology Assessment* 20.63 (ago. 2016), pp. 1–96. DOI: 10.3310/hta20630. URL: <https://doi.org/10.3310/hta20630> (cit. a p. 3).
- [14] J. J. M. Askenasy. «Sleep in Parkinson's disease». In: *Acta Neurologica Scandinavica* 87.3 (gen. 2009), pp. 167–170. DOI: 10.1111/j.1600-0404.1993.tb04095.x. URL: <https://doi.org/10.1111/j.1600-0404.1993.tb04095.x> (cit. a p. 5).
- [15] Ambra Stefani e Birgit Högl. «Sleep in Parkinson's disease». In: *Neuropsychopharmacology* 45.1 (giu. 2019), pp. 121–128. DOI: 10.1038/s41386-019-0448-y. URL: <https://doi.org/10.1038/s41386-019-0448-y> (cit. alle pp. 5, 34).
- [16] J.P.W.F. Lakke, T.W. van Weerden e A. Staal-Schreinemachers. «Axial apraxia, a distinct phenomenon». In: *Clinical Neurology and Neurosurgery* 86.4 (gen. 1984), pp. 291–294. DOI: 10.1016/0303-8467(84)90291-9. URL: [https://doi.org/10.1016/0303-8467\(84\)90291-9](https://doi.org/10.1016/0303-8467(84)90291-9) (cit. a p. 5).
- [17] K R Chaudhuri. «The Parkinson's disease sleep scale: a new instrument for assessing sleep and nocturnal disability in Parkinson's disease». In: *Journal of Neurology, Neurosurgery & Psychiatry* 73.6 (dic. 2002), pp. 629–635. DOI: 10.1136/jnnp.73.6.629. URL: <https://doi.org/10.1136/jnnp.73.6.629> (cit. a p. 5).
- [18] V. Dhawan, D. G. Healy, S. Pal e K. Ray Chaudhuri. «Sleep-related problems of Parkinson's disease». In: *Age and Ageing* 35.3 (mag. 2006), pp. 220–228. DOI: 10.1093/ageing/afj087. URL: <https://doi.org/10.1093/ageing/afj087> (cit. a p. 5).

- [19] Kallol Ray Chaudhuri et al. «International multicenter pilot study of the first comprehensive self-completed nonmotor symptoms questionnaire for Parkinson’s disease: The NMSQuest study». In: *Movement Disorders* 21.7 (mar. 2006), pp. 916–923. DOI: 10.1002/mds.20844. URL: <https://doi.org/10.1002/mds.20844> (cit. a p. 6).
- [20] Derek K. Shaeffer. «MEMS inertial sensors: A tutorial overview». In: *IEEE Communications Magazine* 51.4 (apr. 2013), pp. 100–109. DOI: 10.1109/mcom.2013.6495768. URL: <https://doi.org/10.1109/mcom.2013.6495768> (cit. a p. 7).
- [21] Manon Kok, Jeroen D. Hol e Thomas B. Schön. «Using Inertial Sensors for Position and Orientation Estimation». In: *Foundations and Trends® in Signal Processing* 11.1-2 (2017), pp. 1–153. DOI: 10.1561/20000000094. URL: <https://doi.org/10.1561/20000000094> (cit. a p. 12).
- [22] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. 1st ed. 2006. Corr. 2nd printing. Information science and statistics. Springer, 2006. ISBN: 9780387310732,0387310738 (cit. alle pp. 12, 18, 19).
- [23] Chih-Chung Chang e Chih-Jen Lin. «LIBSVM: A Library for Support Vector Machines». In: *ACM Trans. Intell. Syst. Technol.* 2.3 (mag. 2011). ISSN: 2157-6904. DOI: 10.1145/1961189.1961199 (cit. a p. 15).
- [24] Adele Cutler, D. Richard Cutler e John R. Stevens. «Random Forests». In: *Ensemble Machine Learning*. Springer US, 2012, pp. 157–175. DOI: 10.1007/978-1-4419-9326-7\_5. URL: [https://doi.org/10.1007/978-1-4419-9326-7\\_5](https://doi.org/10.1007/978-1-4419-9326-7_5) (cit. a p. 19).
- [25] Alexey Natekin e Alois Knoll. «Gradient boosting machines, a tutorial». In: *Frontiers in Neurorobotics* 7 (2013). DOI: 10.3389/fnbot.2013.00021. URL: <https://doi.org/10.3389/fnbot.2013.00021> (cit. a p. 20).
- [26] Pratap Dangeti. *Statistics for Machine Learning*. Birmingham, UK: Packt Publishing, 2017 (cit. a p. 22).
- [27] Oluremi A. Famodu, Makenzie L. Barr, Ida Holásková, Wenjun Zhou, Jesse S. Morrell, Sarah E. Colby e Melissa D. Olfert. «Shortening of the Pittsburgh Sleep Quality Index Survey Using Factor Analysis». In: *Sleep Disorders* 2018 (2018), pp. 1–9. DOI: 10.1155/2018/9643937. URL: <https://doi.org/10.1155/2018/9643937> (cit. a p. 26).
- [28] Federica Amato, Irene Rechichi, Luigi Borzi e Gabriella Olmo. «Sleep Quality through Vocal Analysis: a Telemedicine Application». In: *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, mar. 2022. DOI: 10.1109/percomworkshops53856.2022.9767372. URL: <https://doi.org/10.1109/percomworkshops53856.2022.9767372> (cit. alle pp. 28, 65).

- [29] Fei Xue et al. «Analysis of nocturnal hypokinesia and sleep quality in Parkinson's disease». In: *Journal of Clinical Neuroscience* 54 (ago. 2018), pp. 96–101. DOI: 10.1016/j.jocn.2018.06.016. URL: <https://doi.org/10.1016/j.jocn.2018.06.016> (cit. a p. 29).
- [30] Jacob Benesty, Jingdong Chen, Yiteng Huang e Israel Cohen. «Pearson Correlation Coefficient». In: *Noise Reduction in Speech Processing*. Springer Berlin Heidelberg, 2009, pp. 1–4. DOI: 10.1007/978-3-642-00296-0\_5. URL: [https://doi.org/10.1007/978-3-642-00296-0\\_5](https://doi.org/10.1007/978-3-642-00296-0_5) (cit. a p. 37).
- [31] Geng Yang, Zhibo Pang, M. Jamal Deen, Mianxiong Dong, Yuan-Ting Zhang, Nigel Lovell e Amir M. Rahmani. «Homecare Robotic Systems for Healthcare 4.0: Visions and Enabling Technologies». In: *IEEE Journal of Biomedical and Health Informatics* 24.9 (set. 2020), pp. 2535–2549. DOI: 10.1109/jbhi.2020.2990529. URL: <https://doi.org/10.1109/jbhi.2020.2990529> (cit. a p. 58).