

# POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Biomedica



Tesi di Laurea Magistrale

**Supporto alla pratica clinica in pazienti affetti da malattia di Parkinson tramite analisi della voce: confronto tra tecniche di registrazione per l'identificazione di un protocollo efficace per monitoraggio da remoto**

Relatori

Prof.ssa GABRIELLA OLMO

Dott.ssa FEDERICA AMATO

Candidata

MARTINA BENVENUTO

Marzo 2023



# Sommario

La malattia di Parkinson (PD) è una condizione cronica e progressiva causata dal deterioramento neuronale nella substantia nigra, regione del mesencefalo, implicata, tra l'altro, nella produzione di neurotrasmettitori dopaminergici, deputati al controllo motorio. La produzione vocale, implicando una coordinazione complessa e precisa del sistema respiratorio, della laringe e degli articolatori sopraglottali, risulta essere un candidato eccellente per quantificare l'alterazione motoria e fornire informazioni sullo stato di salute del paziente.

Le alterazioni della voce e del linguaggio si verificano in circa il 75-90% della popolazione Parkinson. Si definisce "*disartria*" il disturbo neuromotorio che coinvolge la componente motoria del processo di produzione del linguaggio, e "*disfonia*" l'incapacità di produrre una normale fonazione dovuta al funzionamento alterato del sistema fonatorio.

Le analisi acustiche consentono la quantificazione oggettiva di questi cambiamenti caratteristici nella voce. Questo, insieme al fatto che la registrazione e l'elaborazione del linguaggio umano è un'area con un vasto background di conoscenza, fornisce ai medici una metodologia non invasiva e affidabile per la valutazione dello stato di salute e monitoraggio della malattia.

A questo proposito, la valutazione vocale tramite smartphone possiede intriganti potenzialità in quanto poco costosa, non invasiva, scalabile, semplice da gestire e eseguibile dal paziente. In aggiunta, le registrazioni possono essere inviate ad un server per l'elaborazione presso un centro specializzato o possono essere elaborate direttamente su smartphone mediante analisi acustiche del parlato completamente automatizzate.

Lo scopo che si pone questo lavoro di tesi risulta essere la validazione delle registrazioni vocali acquisite con smartphone, e dunque il confronto delle caratteristiche acustiche maggiormente rappresentative della malattia e invariante rispetto alla modalità di registrazione.

Con questo fine, sono stati utilizzati 2 dataset: (a) il primo composto da 35 controlli, senza alcun sintomo associato al Parkinson, e 64 pazienti Parkinsoniani; (b) il secondo da 15 pazienti Parkinsoniani i cui dati sono stati raccolti presso

*l'Associazione Amici Parkinsoniani Piemonte (AAPP)*. In entrambi i dataset i segnali vocali sono stati registrati contemporaneamente sia con apparecchiatura professionale che con un comune smartphone. Ai fini di questa tesi, sono state analizzate solo le registrazioni provenienti dalla fonazione prolungata della vocale /a/. Il primo dei due dataset è stato utilizzato per la creazione di un algoritmo di rilevamento del Parkinson, il secondo per effettuare analisi statistiche e validare i risultati precedentemente ottenuti.

I risultati complessivi di questa tesi permettono dunque di capire quali parametri, invarianti rispetto alle tecniche di registrazione, possano essere estratti dal segnale vocale ed utilizzati per la detezione e monitoraggio della malattia di Parkinson.

# Ringraziamenti

Ringrazio la mia relatrice, la professoressa Gabriella Olmo, la quale ha dimostrato sin dal primo giorno un grande interesse nei miei confronti, coinvolgendomi attivamente nelle attività con i pazienti. Grazie a lei, ho potuto immergermi appieno nel lavoro di tesi, vivendo questa esperienza con entusiasmo e passione. Vorrei inoltre ringraziarla per la sua grande disponibilità e gentilezza.

Ringrazio la mia correlatrice Federica, che con costanza e dedizione mi ha seguito durante questo meraviglioso percorso, spronandomi a raggiungere obiettivi sempre diversi e supportandomi nei momenti di bisogno.

Grazie ai miei genitori, a voi che mi avete amato incondizionatamente, anche se questo a volte non l'ho compreso. Il vostro sostegno e incoraggiamento mi hanno permesso di crescere e maturare, aiutandomi ad acquisire le competenze e le qualità necessarie per diventare la persona che sono oggi. Vi voglio bene.

Grazie a mia sorella Serena, che sin da piccola nel "lettone con mamma" mi ha insegnato a sorridere. Vorrei avere più tempo da dedicarti, per riuscire a dimostrarti quanto io sia orgogliosa e fiera di te. Credi sempre in te stessa, non sentirti mai sbagliata, e amati infinitamente: la tua felicità è la mia.

Sono immensamente grata a te, nonna Teresa, per il modo in cui hai sempre parlato alla me "bambina". Hai saputo trasmettermi dolcezza, amore e attenzione in ogni momento della mia vita, regalandomi il tuo tempo prezioso e la tua saggezza. Hai saputo insegnarmi il valore della vita e l'importanza di apprezzare ogni piccola cosa, ti voglio bene.

Grazie a nonno Nicola per il bene immenso che mi doni. A te che per primo non vedevi l'ora di chiamarmi "Dottoressa": ora sarai doppiamente orgoglioso di chiamarmi "Dottoressa Bis".

Grazie a nonno Santo e a nonna Francesca, che da lassù hanno sempre fatto il tifo per me.

Sono profondamente grata ai miei bisnonni per avermi insegnato che l'amore è la forza più grande e potente che esiste al mondo. Un abbraccio dalla vostra dottoressa, vi voglio bene.

Grazie *Zi Calì* per essere sempre stata presente nella mia vita, da che ne ho memoria. Grazie anche a zio Alessandro, Nicole e Cristian: siete e sarete la mia seconda famiglia.

Grazie Nicole, perchè con i tuoi abbracci mi fai tornare un pò bambina.

Un grazie infinito va a te, Antonio. Tu più di tutti mi sei stato accanto, nel modo più autentico, dolce e premuroso che si possa immaginare. Giorno dopo giorno è entusiasmante crescere insieme, modellare un futuro tutto nostro con ambizioni e obiettivi comuni. Ti ringrazio di cuore, perché tu mantieni viva la parte più bella di me. A te che hai sempre creduto in me, anche quando io stessa non riuscivo: ti amo.

Ringrazio Mariagrazia, la mia guida sin da piccola. Grazie ai tuoi insegnamenti e alla tua costante fiducia nelle mie capacità, ho capito di voler intraprendere questo percorso. Sei sempre stata un esempio per me e continui ad esserlo tutt'ora. Grazie di cuore per essere stata presente nella mia vita e per avermi sostenuto in ogni mia scelta.

Ringrazio mio cugino Alessandro per essere stato presente nei momenti più belli e importanti della mia vita, in ogni modo possibile.

Un enorme ringraziamento va a te Elena, mia migliore amica. A te che mi hai ascoltata senza giudicarmi, mi hai donato il tuo tempo, i tuoi sorrisi, i tuoi momenti di sana pazzia: sarò sempre legata a te, ovunque la nostra vita ci porterà. Con te, il tempo sembra non passare mai, rimanendo fermo tra i banchi di scuola o in quei pomeriggi passati a confidarci le nostre paure e gioie. Ti voglio bene Elenù.

Ringrazio Roberta, Fabiana, Lucia, Luca, Gianpaolo, Martina e Elena: con voi mi sento libera di essere me stessa. Ogni volta vorrei che il tempo si fermasse ai nostri brindisi, alle nostre risate, ai nostri balli scatenati. Siete la mia famiglia, quella che ho scelto, e non potrei chiedere di meglio. Dal profondo del mio cuore, vi voglio bene.

Grazie di cuore Gianpaolo, sei stato una bellissima scoperta nella mia vita. Grazie per il tuo sorriso contagioso, la tua sincerità: quando ci sei tu, il tempo vola. E ricorda, mi casa es tu casa.

Viola, Elisabetta e Sharon, non potrei mai ringraziarvi abbastanza per aver reso "Via Gioberti 69" un posto magico, un porto sicuro, un luogo dove ho trovato la mia serenità. Siete state la mia famiglia lontano dalla mia famiglia. Con tutto il cuore, vi voglio bene.

Grazie a Emanuele, Martina, Lucia, Frank e Cristina per aver reso questi anni al Politecnico indimenticabili. Quell'aula studio è diventata un simbolo delle nostre lunghe ore di studio e del sostegno reciproco che ci siamo dati durante questi anni:

ho lasciato un pezzo di cuore tra quei banchi. A voi che avete reso il Poli una seconda casa, vi voglio bene.

Grazie Matteo, che dal primo giorno a Garda, mi hai conquistata con la tua simpatia, che solo un calabrese possiede. Grazie a quei giorni trascorsi insieme, ho avuto la fortuna di incontrare una persona speciale come te, un vero amico su cui poter contare in qualsiasi momento.

"Sei brava, non lo sai, ma lo sei, sei la mia spalla": infinitamente grazie Benni di essere stata la mia spalla in questi anni.

Carla bedda, non potrei essere più grata per la cura e la gentilezza che mi dimostri costantemente. Sei una ragazza dal cuore d'oro e sono felice di averti nella mia vita.

Desidero ringraziare di cuore tutti i miei zii, parenti, familiari e amici, che in questi anni mi hanno offerto il loro prezioso tempo, la loro compagnia, gli abbracci, le belle parole e l'amore. Non avrei potuto arrivare fin qui senza il vostro sostegno e il vostro affetto. Vi porto tutti con me, sempre.

Il mio ringraziamento più sentito va alla bimba che è in me. Sei stata tu, fin dall'inizio, a guidare ogni mia decisione, a farmi vedere la vita con occhi curiosi e a darmi la forza di sognare in grande. Grazie per avermi insegnato a guardare al mondo con semplicità e meraviglia, e per avermi dato la forza di credere in me stessa e nei miei sogni. Continuerò a seguirti ovunque tu vada, perché sei la parte più autentica e speciale di me.

A Torino, la mia città, il mio rifugio, la mia ispirazione. Sarò per sempre legata indissolubilmente alla tua bellezza. Grazie.





# Indice

<b>Elenco delle tabelle</b>	XI
<b>Elenco delle figure</b>	XIV
<b>1 Introduzione</b>	1
1.1 Apparato vocale e fonazione . . . . .	1
1.2 Malattia di Parkinson e fonazione . . . . .	2
<b>2 Stato dell'arte</b>	5
2.1 Task di registrazione . . . . .	5
2.2 Detezione automatica del PD . . . . .	6
2.3 Smartphone e microfono a confronto . . . . .	7
<b>3 Materiali e Metodi</b>	9
3.1 Partecipanti e procedure sperimentali . . . . .	9
3.2 Acquisizione dati . . . . .	11
3.2.1 Dataset LUHS . . . . .	11
3.2.2 Dataset AAPP . . . . .	11
3.3 Features vocali . . . . .	12
3.3.1 OpenSMILE features . . . . .	16
3.3.2 Essentia descriptors . . . . .	17
3.3.3 jAudio features . . . . .	17
3.3.4 MPEG7 descriptors . . . . .	19
3.3.5 Tsanas features . . . . .	19
3.3.6 YAAFE features . . . . .	19
3.4 Classificazione PD . . . . .	19
3.4.1 Preparazione dei dati e dei segnali . . . . .	20
3.4.2 Feature Selection . . . . .	21
3.4.3 Classificazione . . . . .	22
3.4.4 Validazione incrociata . . . . .	27
3.5 Analisi statistica . . . . .	28

<b>4</b>	<b>Risultati</b>	<b>33</b>
4.1	Classificazione PD . . . . .	33
4.1.1	Validazione incrociata . . . . .	44
4.2	Analisi statistica . . . . .	44
<b>5</b>	<b>Discussioni</b>	<b>53</b>
5.1	Classificazione PD . . . . .	53
5.1.1	Validazione incrociata . . . . .	56
5.2	Analisi statistica . . . . .	57
<b>6</b>	<b>Conclusioni</b>	<b>65</b>
	<b>Bibliografia</b>	<b>67</b>

# Elenco delle tabelle

3.1	Tabella riassuntiva dei soggetti del dataset LUHS. AC: microfono professionale; SP: smartphone; CS: Soggetti Controllo Sani; PD: Soggetti affetti da Parkinson. . . . .	10
3.2	Tabella riassuntiva dei soggetti del dataset AAPP. AC: microfono professionale; SP: smartphone. . . . .	10
3.3	Tabella riassuntiva set di features del dataset LUHS. . . . .	16
3.4	Tabella riassuntiva features ComPare [52]. . . . .	17
3.5	Tabella riassuntiva delle famiglie di features Essentia [53]. . . . .	18
4.1	Valori di Validation Accuracy massima per ogni subset al variare della Feature Selection utilizzata, per registrazioni AC. Si evidenziano i valori maggiori di Validation Accuracy per ogni subset. . . . .	34
4.2	Valori di Validation Accuracy massima per ogni subset al variare della Feature Selection utilizzata, per registrazioni SP. Si evidenziano i valori maggiori di Validation Accuracy per ogni subset. . . . .	35
4.3	Valori di validation accuracy (AC) al variare del numero di features da selezionare dell’algoritmo <i>ANOVA</i> . Si riporta il numero del subset considerato (No) e il classificatore con cui è stato raggiunto tale risultato. Si evidenziano i valori di Validation accuracy migliori. . .	36
4.4	Valori di validation accuracy (AC) al variare del valore della soglia della correlazione di Pearson $r$ dell’algoritmo <i>Correlation-based approach</i> . Si riporta il numero del subset considerato (No) e il classificatore con cui è stato raggiunto tale risultato. Si evidenziano i valori di Validation accuracy migliori. . . . .	36
4.5	Valori di validation accuracy (AC) al variare del valore del parametro percentuale dell’algoritmo <i>BORUTA</i> . Si riporta il numero del subset considerato (No) e il classificatore con cui è stato raggiunto tale risultato. Si evidenziano i valori di Validation accuracy migliori. . .	37

4.6	Valori di validation accuracy (SP) al variare del numero di features da selezionare dell'algoritmo <i>ANOVA</i> . Si riporta il numero del subset considerato (No) e il classificatore con cui è stato raggiunto tale risultato. Si evidenziano i valori di Validation accuracy migliori. . . . .	37
4.7	Valori di validation accuracy (SP) al variare del valore della soglia della correlazione di Pearson <i>r</i> dell'algoritmo <i>Correlation-based approach</i> . Si riporta il numero del subset considerato (No) e il classificatore con cui è stato raggiunto tale risultato. Si evidenziano i valori di Validation accuracy migliori. . . . .	38
4.8	Valori di validation accuracy (SP) al variare del valore del parametro percentuale dell'algoritmo <i>BORUTA</i> . Si riporta il numero del subset considerato (No) e il classificatore con cui è stato raggiunto tale risultato. Si evidenziano i valori di Validation accuracy migliori. . . . .	38
4.9	Valori di validation accuracy e test accuracy (AC) per le coppie migliori Feature Selection - Classificatore. . . . .	41
4.10	Valori di validation accuracy e test accuracy (SP) per le coppie migliori Feature Selection - Classificatore. . . . .	42
4.11	Migliore coppia FS-Classificatore (AC e SP): si riportano i relativi subset e le feature selezionate. . . . .	43
4.12	Valori di train accuracy, test accuracy e test accuracy in seguito a validazione incrociata (Test accuracy Val.Incr): lungo le colonne vengono specificati i subset di partenza, mentre lungo le righe si descrivono le tecniche utilizzate di classificazioni e le metriche di performance calcolate. . . . .	44
5.1	AC: Valori a confronto di test accuracy tra l'algoritmo di classificazione sviluppato nel lavoro di tesi e quello dello studio [14]. Si evidenziano i set di dati con cui si raggiungono valori più alti nel lavoro di tesi rispetto allo studio. FS_T: Feature Selection utilizzata nel lavoro di tesi; FS_A: Feature Selection utilizzata nello studio; Class_T: Classificatore utilizzato nel lavoro di tesi; Class_A: Classificatore utilizzato nello studio; Test_Accuracy_T: Accuratezza raggiunta nel lavoro di tesi; Test_Accuracy_A: Accuratezza raggiunta nello studio. . . . .	54

5.2	SP: Valori a confronto di test accuracy tra l'algoritmo di classificazione sviluppato nel lavoro di tesi e quello dello studio [14]. Si evidenziano i set di dati con cui si raggiungono valori più alti nel lavoro di tesi rispetto allo studio. FS_T: Feature Selection utilizzata nel lavoro di tesi; FS_A: Feature Selection utilizzata nello studio; Class_T: Classificatore utilizzato nel lavoro di tesi; Class_A: Classificatore utilizzato nello studio; Test_Accuracy_T: Accuratezza raggiunta nel lavoro di tesi; Test_Accuracy_A: Accuratezza raggiunta nello studio. . . . .	55
5.3	Valori a confronto di test accuracy ottenuti con registrazioni AC e SP. Si evidenziano i valori di accuratezza più alti. . . . .	56

# Elenco delle figure

1.1	Apparato fonatorio umano . . . . .	2
3.1	Diagramma schematico del set-up sperimentale per la registrazione del segnale vocale con 2 device diversi simultaneamente. . . . .	11
3.2	Misura di jitter in un segnale vocale [8]. . . . .	13
3.3	Misura di shimmer in un segnale vocale [8]. . . . .	14
3.4	Misura di HNR in un segnale vocale [8]. . . . .	15
3.5	Flowchart algoritmo classificazione . . . . .	20
3.6	Funzionamento di base del Support Vector Machine classifier [61]. . . . .	23
3.7	Funzionamento di base del K-Nearest Neighbors classifier, con un valore di $k=1$ . . . . .	24
3.8	Flowchart nel dettaglio dell'algoritmo di classificazione. . . . .	25
3.9	Flowchart nel dettaglio dell'ottimizzazione della Feature Selection, a partire dai risultati ottenuti dalla Classificazione. . . . .	26
3.10	Descrizione utile alla comprensione dei dati utilizzati per la validazione incrociata. . . . .	27
3.11	Schema della procedura di validazione incrociata. . . . .	28
4.1	Migliore coppia Feature Selection - Classificatore per ogni subset (AC): si riportano i valori dei parametri ottimizzati. . . . .	39
4.2	Migliore coppia Feature Selection - Classificatore per ogni subset (SP): si riportano i valori dei parametri ottimizzati. . . . .	40
4.5	Elenco delle prime 20 feature (subset 13) con p-value $< 0.05$ : vengono riportate a destra le famiglie di features a cui appartengono. . . . .	45
4.3	Il grafico riporta il numero di volte in cui una feature con p-value $< 0.05$ figura tra le prime 20 feature, ordinate in funzione della significatività restituita dal test di Wilcoxon, per ciascuno dei subset di dati in LUHS. La dimensione di ciascun segmento rappresenta il numero di feature relativo alla famiglia di feature e al toolbox considerato. . . . .	46

4.4	Il grafico riporta il numero di volte in cui una feature con p-value $> 0.05$ figura tra le ultime 20 feature, ordinate in funzione della significatività restituita dal test di Wilcoxon, per ciascuno dei subset di dati in LUHS. La dimensione di ciascun segmento rappresenta il numero di feature relativo alla famiglia di feature e al toolbox considerato. . . . .	47
4.6	Boxplot delle distribuzioni di valori delle features del dataset AAPP appartenenti alle registrazioni AC, SP e SP denoised (continua). . .	48
4.6	Boxplot delle distribuzioni di valori delle features del dataset AAPP appartenenti alle registrazioni AC, SP e SP denoised (continua). . .	49
4.6	Boxplot delle distribuzioni di valori delle features del dataset AAPP appartenenti alle registrazioni AC, SP e SP denoised (continua). . .	50
4.6	Boxplot delle distribuzioni di valori delle features del dataset AAPP appartenenti alle registrazioni AC, SP e SP denoised (continua). . .	51
4.6	Boxplot delle distribuzioni di valori delle features del dataset AAPP appartenenti alle registrazioni AC, SP e SP denoised. . . . .	52
5.1	Il grafico riporta il numero di volte in cui i descrittori delle feature <i>Spectral</i> con p-value $< 0.05$ figurano tra le prime 20 feature, ordinate in funzione della significatività restituita dal test di Wilcoxon, per ciascuno dei subset di dati in LUHS. . . . .	60
5.2	Il grafico riporta il numero di volte in cui i descrittori delle feature <i>Spectral</i> con p-value $> 0.05$ figurano tra le ultime 20 feature, ordinate in funzione della significatività restituita dal test di Wilcoxon, per ciascuno dei subset di dati in LUHS. . . . .	61
5.3	Il grafico riporta il numero di volte in cui i descrittori delle feature <i>MFCC</i> con p-value $< 0.05$ figurano tra le prime 20 feature, ordinate in funzione della significatività restituita dal test di Wilcoxon, per ciascuno dei subset di dati in LUHS. . . . .	62
5.4	Il grafico riporta il numero di volte in cui i descrittori delle feature <i>MFCC</i> con p-value $> 0.05$ figurano tra le ultime 20 feature, ordinate in funzione della significatività restituita dal test di Wilcoxon, per ciascuno dei subset di dati in LUHS. . . . .	63





# Capitolo 1

## Introduzione

### 1.1 Apparato vocale e fonazione

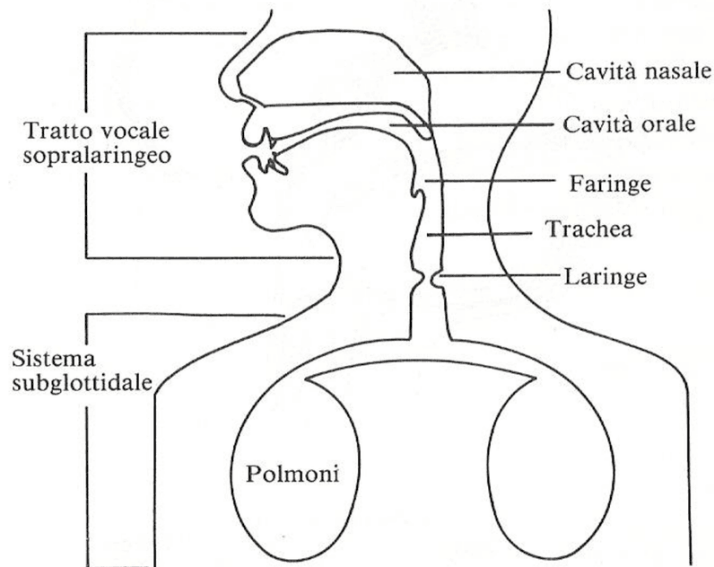
Da un punto di vista ingegneristico è possibile modellizzare l'apparato fonatorio come un sistema fluido-meccanico [1] costituito da:

- polmoni, che fungono da generatore di pressione continua;
- corde vocali, ovvero l'attuatore;
- tratto vocale, il risuonatore.

Il processo di produzione del segnale vocale origina dai polmoni i quali generano un flusso d'aria che viene dapprima forzato verso il tratto bronchiale e successivamente attraverso l'area glottale. In corrispondenza dell'area glottale sono localizzate le corde vocali (o pliche vocali), strutture di tessuto fortemente elastico caratterizzato da elevata capacità di contrazione e rilassamento. Il suono "grezzo" che si origina dal passaggio del flusso d'aria attraverso le corde vocali, e l'eventuale messa in vibrazione di queste ultime [2], è l'elemento costituente della voce di ciascun individuo. Nel caso in cui le corde vocali si trovino in posizione aperta e consentano il passaggio del flusso d'aria senza ostacolarne il percorso, il suono prodotto sarà definito *non vocalizzato*. Viceversa, nel caso in cui le corde vocali siano in posizione ravvicinata, costituiranno un ostacolo al passaggio del flusso d'aria proveniente dai polmoni, ed entreranno in vibrazione spinte dalla pressione sub-glottale generatasi. La messa in vibrazione delle corde vocali al passaggio del flusso d'aria produrrà un segnale periodico, la cui frequenza fondamentale è rappresentata dalla frequenza di oscillazione delle corde vocali. Come in qualunque sistema meccanico, tale frequenza sarà influenzata dalle caratteristiche fisiche delle corde, quali lunghezza, tensione e massa con conseguenti differenze sul suono prodotto. Ne risulta che, nei soggetti di sesso maschile, le cui corde vocali hanno

maggiore lunghezza e rigidità, la frequenza fondamentale sia più bassa rispetto ai soggetti di sesso femminile, con un conseguente suono percepito più grave.

Il flusso d'aria generato, sia esso periodico o aperiodico, verrà poi ulteriormente modulato da altre strutture nel tratto fonatorio, ovvero labbra, denti e lingua, generalmente definiti *articolatori*. Nella Figura 1.1 viene mostrato questo sistema.



**Figura 1.1:** Apparato fonatorio umano

## 1.2 Malattia di Parkinson e fonazione

La malattia di Parkinson (PD) è una condizione cronica e progressiva causata dal graduale deterioramento neuronale nella substantia nigra, implicata nella produzione di neurotrasmettitori dopaminergici, che svolgono un ruolo cruciale nel controllo motorio.

A causa di tale deterioramento, la malattia di Parkinson risulta essere caratterizzata da una grande varietà di sintomi motori, che comprendono tremore, movimenti lenti (bradicinesia), rigidità e instabilità posturale. In aggiunta i pazienti sono soliti riportare sintomi non motori come disturbi cognitivi/neurocomportamentali, anomalie sensoriali e del sonno [3]. Possono anche insorgere negli stadi avanzati della malattia problemi comportamentali, depressione, ansia e demenza [4].

La diagnosi della malattia di Parkinson viene effettuata sulla base di criteri clinici e dunque in seguito ad un esame neurologico dettagliato, comprensivo di una revisione dell'anamnesi del paziente e di una valutazione clinica dei sintomi

motori e non. Il monitoraggio della progressione della malattia viene generalmente eseguito alcune volte all'anno durante visite ambulatoriali. Attualmente la scala MDS-UPDRS (Movement Disorder Society revision of the Unified Parkinson's Disease Rating Scale) è universalmente impiegata per valutare il decorso della malattia dopo la sua diagnosi [3].

La fonazione, implicando una coordinazione complessa e precisa del sistema respiratorio, della laringe e degli articolatori sopraglottali, risulta essere un candidato eccellente per fornire informazioni diagnostiche [5]. Le alterazioni della voce e del linguaggio si verificano in circa il 75-90% della popolazione Parkinson [6]. Più in dettaglio, la *disartria* è un disturbo neuromotorio che coinvolge la componente motoria del processo di produzione del linguaggio ed è correlato a limitazioni respiratorie, allungamento o adduzione ridotti delle corde vocali e disturbi a livello articolatorio. È caratterizzato da una scarsa articolazione dei fonemi (alterazioni di forza, velocità, volume, tono, gamma o precisione dei movimenti necessari per il controllo vocale [7]), con comprensione e ideazione del linguaggio intatte. La *disfonia* può essere invece definita come l'incapacità di produrre una normale fonazione dovuta al funzionamento alterato del sistema fonatorio, mentre la disartria è più legata ai problemi di articolazione durante la pronuncia delle parole [5].

Le analisi acustiche consentono la quantificazione di questi cambiamenti caratteristici nella voce. Questo, insieme al fatto che la registrazione e l'elaborazione del linguaggio umano è un'area con un vasto background di conoscenza, rende il cambiamento della fonazione un eccellente candidato come biomarcatore diagnostico del PD [8]; inoltre fornisce ai medici una metodologia non invasiva e affidabile per la diagnosi, valutazione e monitoraggio della malattia [9].

A questo proposito, la valutazione vocale tramite smartphone possiede intriganti potenziali in quanto poco costosa, non invasiva, scalabile, semplice da gestire. Essa può essere eseguita a distanza dal domicilio del paziente. In aggiunta, le registrazioni possono essere inviate ad un server per l'elaborazione presso un centro specializzato da remoto o possono essere elaborate direttamente su smartphone mediante analisi acustiche del parlato completamente automatizzate [10]. Dato il progresso nello sviluppo di tecnologie mobili, la raccolta di dati attraverso gli smartphone continua a essere un focus crescente nell'ambito della ricerca E-health. Gli smartphone, infatti, grazie alla loro ubiquità, consentono una raccolta di dati molto più frequente con costo inferiore.

E' però risaputo che le prestazioni risultanti dall'analisi delle caratteristiche acustiche possono differire in base ai vari dispositivi di registrazione e microfoni utilizzati [11]. Tuttavia, in base a quanto emerso dall'analisi della letteratura, non si è ancora concordi circa l'entità e la tipologia dell'influenza che la modalità di acquisizione dei campioni audio abbia sui modelli di AI (Intelligenza artificiale) realizzati. In questo contesto, lo scopo che si pone questo lavoro di tesi risulta

essere la validazione delle registrazioni vocali acquisite con smartphone, e dunque il confronto delle caratteristiche acustiche rappresentative ottenute parallelamente attraverso uno smartphone e un microfono professionale.

# Capitolo 2

## Stato dell'arte

### 2.1 Task di registrazione

Negli ultimi decenni si è visto un crescente interesse della comunità scientifica verso lo studio dell'analisi vocale per supportare la pratica clinica nel monitoraggio del PD. In generale è possibile effettuare una suddivisione ad alto livello di tali studi basandosi sul task di registrazione scelto e, di conseguenza, sulla dimensione vocale analizzata. In generale è possibile dividerli in quattro gruppi principali a seconda dell'aspetto vocale analizzato, cioè fonatorio, articolatorio, prosodico e cognitivo-linguistico [5]:

- Gli studi *fonatori* sono correlati alla sorgente glottale e alle strutture di risonanza del tratto vocale. Impiegano le *vocali sostenute* come materiale acustico per l'analisi.
- Gli studi basati sugli aspetti *articolatori* sono più diversificati in quanto esistono maggiori possibilità di analisi. Le caratteristiche o le misure acustiche analizzate possono essere estratte da diversi tipi di segmenti sonori e possono essere correlate alla velocità, all'accelerazione degli articolatori o alla tipologia di transizione tra i segmenti.
- Gli studi *prosodici* si concentrano principalmente su caratteristiche paralinguistiche come variazione di tono, analisi del ritmo delle sillabe o manifestazione di emozioni nel segnale vocale.
- Gli approcci *cognitivo-linguistici* analizzano le deviazioni in comportamento cognitivo esaminando il vocabolario, la complessità delle frasi e l'esistenza di ripetizioni di parole.

Tanto il compito vocale quanto il materiale acustico utilizzato in ciascun caso sono fattori chiave di differenziazione dei quattro gruppi. Nell'analisi fonatoria, le *vocali*

*sostenute* sono comunemente impiegate come materiale acustico, mentre negli altri tre gruppi è necessario l'uso del *discorso connesso* (monologhi, lettura di brani).

Gli studi fonatori sono i più diffusi, dal momento che fanno uso di un task molto semplice e di facile analisi: al paziente viene chiesto di produrre una vocale singola e mantenere il tono di questa il più costante e più a lungo possibile. Sebbene un discorso connesso possa essere considerato come un test più realistico di verifica della compromissione vocale, da semplici test di fonazione sostenuta si è in grado di ricavare sintomi disfonici [12]. Inoltre i test per la rilevazione della disfonia sono condotti al meglio senza gli effetti confondenti dell'articolazione o delle componenti linguistiche del discorso connesso [12]. Il vantaggio delle fonazioni sostenute risiede tanto nella validità delle informazioni fornite su tremore e rigidità delle strutture respiratorie e glottali, quanto nella semplicità del task, poichè non è richiesto di leggere o memorizzare frasi. Inoltre, questi modelli di rilevamento possono essere considerati minimamente influenzati dalla lingua.

Tuttavia, i segnali di fonazione sostenuta sono meno complessi di quelli provenienti da un discorso connesso, per cui questo risulta essere più adatto per analisi degli aspetti cinetici [13].

## 2.2 Detezione automatica del PD

Recentemente, il riconoscimento automatico della malattia di Parkinson attraverso i dati vocali ha richiamato notevole attenzione per i seguenti motivi:

- i disturbi vocali risultano essere tra i primi sintomi della malattia [14];
- quasi il 90% delle persone affette da Parkinson mostra disturbi della voce [14];
- permette quella che viene definita *telediagnosi* della malattia [14];

In letteratura sono molti gli articoli riguardanti il rilevamento automatico del Parkinson: le prestazioni di modelli di classificazione (comunemente valutati in termini di accuratezza) che sfruttano l'analisi vocale variano dal 65 al 99% per stadi della malattia da moderati ad avanzati [15, 16, 17, 18, 19]. Sono state esplorate diverse metodologie di classificazione per rilevare la malattia utilizzando l'analisi vocale.

In [20] gli autori hanno sfruttato le caratteristiche globali del segnale vocale, come il numero di pause, la deviazione standard (SD) del tono e dell'intensità, insieme a perturbazioni medie di basso livello, come shimmer, jitter, rapporto segnale/rumore, formanti.

In generale, la pipeline dei modelli sviluppati è costituita da un primo step di *Feature Selection* e un secondo step di *classificazione*. Più in dettaglio si esegue una *Feature Selection* che permetta di tenere le feature statisticamente

significative rimuovendo le ridondanze. Infine, le feature selezionate vengono fornite ai *classificatori*, come Support Vector Machine (SVM) [21, 22, 23], k-nearest neighbors (KNN) [23, 24], alberi decisionali (DT)[25], multilayer perceptrons (MLP) [21], o reti neurali (ANN) [26].

Accanto alle feature più classiche, una vasta porzione della letteratura sfrutta i Mel-Frequency Cepstral Coefficient (MFCCs) [27]: queste feature caratterizzano l'inviluppo spettrale e riflettono la forma del tratto vocale. Diversi muscoli e articolatori, come i muscoli costrittori faringei, la lingua, le labbra, la mascella, la laringe, il palato molle e la posizione della laringe modulano le forme e i volumi; inoltre l'accoppiamento di queste cavità conferisce un timbro specifico al suono [27]. Ciò si traduce in un particolare inviluppo spettrale descritto dagli MFCC. Pertanto, i coefficienti estratti dal segnale vocale permettono la quantificazione di informazioni relative tanto ad articolazione quanto a fonazione [27]. Sono molti gli studi ad utilizzare questa tipologia di features per il rilevamento della malattia [28, 29, 30] o il monitoraggio [31, 32].

Alcuni studi hanno esplorato la possibilità del supporto della pratica clinica del PD utilizzando le registrazioni effettuate con applicazioni per smartphone o tablet in alta qualità (frequenza di campionamento di 44 o 48kHz), inviandole successivamente ad un server remoto [8, 33, 34, 35]. Uno degli studi più recenti in merito [14] utilizza un set di registrazioni acquisite sia tramite smartphone che microfono professionale, provenienti da due task vocali differenti, *phonation e speech* [34]. Per tale scopo sono stati utilizzati la fonazione sostenuta della vocale /a/ a un livello di tono e volume confortevole e la pronuncia di una frase foneticamente bilanciata in lingua lituana. Successivamente le feature estratte da ciascun segnale vocale sono state inserite in 5 diversi modelli di classificazione (Linear discriminant analysis LDA, Gaussian Naive Bayes GNB, KNN, SVM, e ANN), sia per le registrazioni acquisite tramite smartphone che tramite microfono professionale, e i valori di *accuratezza* raggiunti per il miglior classificatore, risultano essere:

- 74.66% per le registrazioni acquisite tramite smartphone, utilizzando il modello *GNB*;
- 84.00% per le registrazioni acquisite tramite microfono con modello *LDA*.

## 2.3 Smartphone e microfono a confronto

In diversi studi si è analizzato il segnale vocale nel PD utilizzando registrazioni ottenute tramite smartphone [32, 36, 37, 8] ma pochi risultano essere quelli incentrati sul confronto tra le due possibili modalità di acquisizione: *smartphone e microfono professionale*.

Nello studio [38] il segnale è stato acquisito da volontari senza patologie vocali, sia tramite smartphone che tramite microfono professionale; da queste registrazioni, attraverso il software *Praat* [39], sono stati estratti i parametri principali per la valutazione dell'alterazione vocale, come la *frequenza fondamentale (F0)*, *jitter*, *shimmer*, *HNR (Harmonic to noise ratio)*. Lo scopo di tale studio risiede nello stabilire se i valori dei parametri estratti, ottenuti dalla registrazione della vocale sostenuta /a/ usando vari smartphone, possano essere paragonabili a quelli ricavati da un microfono professionale. Per validare tale ipotesi è stata portata avanti un'analisi statistica attraverso un t-test a campioni indipendenti: solo *F0* e *jitter* non hanno mostrato differenze significative tra le due modalità di acquisizione. Analogamente, i risultati riportati in [40, 41] non hanno mostrato differenze significative nei valori di F0 misurati sulle due tipologie di registrazioni. Infine, per quanto riguarda le altre tipologie di feature, l'analisi della letteratura condotta non ha messo in luce risultati concordi circa l'effetto della strumentazione sui singoli parametri acustici.



# Capitolo 3

## Materiali e Metodi

### 3.1 Partecipanti e procedure sperimentali

I dati utilizzati per il lavoro di tesi sono stati raccolti in parte presso il dipartimento di Otorinolaringoiatria della *Lithuanian University of Health Sciences* (LUHS) e in parte, in prima persona, presso l'*Associazione Amici Parkinsoniani Piemonte* (AAPP). Il primo dataset (LUHS) è stato utilizzato nel lavoro di tesi per la definizione di un algoritmo di rilevamento automatico del PD mentre il secondo (AAPP) con lo scopo di effettuare un confronto tra smartphone e microfono professionale eseguendo delle analisi statistiche.

Più nello specifico, il dataset utilizzato in questo lavoro di tesi per la definizione di un algoritmo di rilevamento della malattia di Parkinson (LUHS) comprende 99 partecipanti, di cui 35 controlli senza alcun sintomo associato al Parkinson e 64 Parkinsoniani (Tabella 3.1). E' stato chiesto ai partecipanti di eseguire i seguenti task:

- ripetere tre volte la vocale sostenuta /a/;
- pronunciare una frase foneticamente bilanciata in lingua lituana.

**Tabella 3.1:** Tabella riassuntiva dei soggetti del dataset LUHS. AC: microfono professionale; SP: smartphone; CS: Soggetti Controllo Sani; PD: Soggetti affetti da Parkinson.

	Fonazione		Discorso connesso	
	AC	SP	AC	SP
CS maschi	11(33)	11(33)	11	11
CS femmine	24(72)	24(72)	24	24
CS totali	35(105)	35(105)	35	35
PD maschi	30(89)	30(90)	29	30
PD femmine	34(101)	34(102)	34	34
PD totali	64(190)	64(192)	63	64
Totali	99(295)	99(297)	98	99
Età media CS (Dev. St.)	41.74 (17.11)			
Età media PD (Dev. St)	64.95 (9.56)			

Passando al secondo dataset (AAPP), questo risulta essere composto da registrazioni di 15 pazienti Parkinsoniani (Tabella 3.2). In merito alle procedure di acquisizione, è stato chiesto ai partecipanti di eseguire i seguenti task:

- ripetere due volte la vocale sostenuta /a/;
- ripetere due volte la vocale sostenuta /e/;
- ripetere due volte il proverbio *"A caval donato non si guarda in bocca"*;
- ripetere due volte il proverbio *"Meglio soli che male accompagnati"*.

**Tabella 3.2:** Tabella riassuntiva dei soggetti del dataset AAPP. AC: microfono professionale; SP: smartphone.

Composizione dataset AAPP					
Modalità	Totale	Maschi	Femmine	Età media	Dev.st. età
AC/SP	15	11	4	70.4	9.19

In questo lavoro di tesi è stato analizzato solo il task della vocale sostenuta.

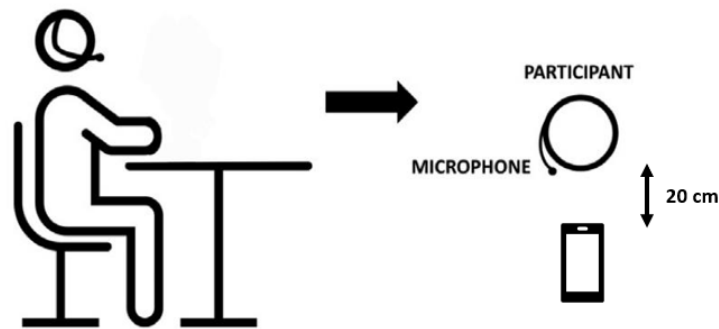
## 3.2 Acquisizione dati

### 3.2.1 Dataset LUHS

I campioni audio sono stati registrati utilizzando due canali contemporaneamente: cardioide acustico (AKG Perception 220, gamma di frequenza 20–20000 Hz) e uno smartphone (microfono interno del Samsung Galaxy Note 3). Entrambi i microfoni erano posizionati a circa 10 cm di distanza dalla bocca. Il formato audio è mono PCM wav (16 bit a una frequenza di campionamento di 44.1 kHz).

### 3.2.2 Dataset AAPP

Per l’acquisizione dei segnali sono stati utilizzati un microfono ad archetto (Microfono Shure) e uno smartphone (iPhone 12) in un ambiente non rumoroso con lo speaker posizionato al centro della stanza per evitare il riverbero del suono dovuto alla vicinanza con le pareti. Inoltre, si è posta attenzione alla posizione della schiena, appoggiata alla sedia così da evitare la compressione della zona diaframmatica, e delle braccia, poggiate sui braccioli della sedia. Il microfono è stato connesso ad un registratore portatile (Zoom h5) che campiona il segnale a una frequenza di 44.1 KHz usando un convertitore analogico-digitale caratterizzato da 16 bit di risoluzione. Allo stesso modo, le registrazioni tramite smartphone, posizionato ad una distanza di circa 20 cm dalla bocca [42], sono state acquisite settando i medesimi parametri.



**Figura 3.1:** Diagramma schematico del set-up sperimentale per la registrazione del segnale vocale con 2 device diversi simultaneamente.

### 3.3 Features vocali

Le famiglie di feature maggiormente utilizzate per analisi vocale di malattie neurodegenerative, come appunto il PD [43], risultano essere le seguenti:

- Frequenza fondamentale (F0 o pitch): viene misurata in Hertz (Hz) e corrisponde al numero di aperture e chiusure della glottide in un periodo di 1 s. Questo parametro è associato al tono di voce di ogni persona e varia in base al sesso e all'età [44]. L'irrigidimento della muscolatura laringea che si verifica con il PD, potrebbe portare un aumento del valore di F0.
- Jitter: è una misura della disfonia influenzata principalmente dalla mancanza di controllo della vibrazione delle corde vocali; le voci patologiche tendono ad avere una percentuale di jitter maggiore rispetto alle voci sane [45]. Questa misura della raucedine mira a quantificare i disturbi da ciclo a ciclo della F0, rappresentati come piccole deviazioni dalla periodicità esatta (Figura 3.2). Per calcolare il jitter viene utilizzato l'inviluppo [46]. Le varianti di jitter sono le seguenti:

- Jitter assoluto ( $\mu\text{s}$ ): misura della variazione del periodo fondamentale da ciclo a ciclo. Si calcola come:

$$Jitter\_abs = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_0^i - T_0^{i+1}|$$

dove  $T_0^i$ ,  $i = 1, 2, \dots, N$  sono i periodi stimati,  $N$  è il numero di periodi stimati.

- Jitter relativo (%): differenza media assoluta tra periodi consecutivi, diviso il periodo medio, come nella seguente formula:

$$Jitter\_rel = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_0^i - T_0^{i+1}|}{\frac{1}{N} \sum_{i=1}^N T_0^i}$$

dove  $T_0^i$ ,  $i = 1, 2, \dots, N$  sono i periodi stimati,  $N$  è il numero di periodi stimati.

- RAP (%): misura relativa della variazione del pitch da periodo a periodo con una media lungo 3 periodi successivi. Si calcola come:

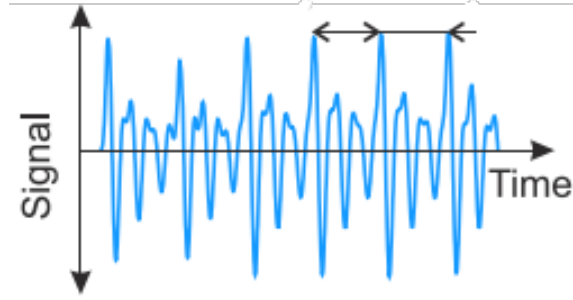
$$RAP = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| \frac{T_0^{i-1} + T_0^i + T_0^{i+1}}{3} - T_0^i \right|}{\frac{1}{N} \sum_{i=1}^N T_0^i}$$

dove  $T_0^i$ ,  $i = 1, 2, \dots, N$  sono i periodi stimati,  $N$  è il numero di periodi stimati.

- PPQ5 (%): misura relativa della variazione del pitch da periodo a periodo con una media lungo 5 periodi successivi. Si calcola come:

$$PPQ5 = \frac{\frac{1}{N-4} \sum_{i=3}^{N-2} \left| \frac{T_0^{i-2} + T_0^{i-1} + T_0^i + T_0^{i+1} + T_0^{i+2}}{5} - T_0^i \right|}{\frac{1}{N} \sum_{i=1}^N T_0^i}$$

dove  $T_0^i$ ,  $i = 1, 2, \dots, N$  sono i periodi stimati,  $N$  è il numero di periodi stimati.



**Figura 3.2:** Misura di jitter in un segnale vocale [8].

- Shimmer: è una misura della disfonia, correlata alla diminuzione della resistenza glottale e a lesioni delle corde vocali [44]. Lo shimmer misura i disturbi dell'ampiezza del segnale vocale (Figura 3.3); si utilizza per i suoi calcoli l'involuppo. L'involuppo è definito utilizzando il valore di ampiezza massima all'interno di ciascun ciclo glottale [46]. Le varianti di shimmer sono le seguenti:

- Shimmer assoluto (dB): misura della variazione dell'ampiezza picco - picco da periodo a periodo. Si calcola come:

$$Shimmer\_dB = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| 20 \log \left( \frac{A^{i+1}}{A^i} \right) \right|$$

dove  $A^i$ ,  $i = 1, 2, \dots, N$  sono le ampiezze picco-picco,  $N$  è il numero di impulsi stimati.

- Shimmer relativo (%): misura relativa della variazione dell'ampiezza picco-picco da periodo a periodo. Si calcola come:

$$Shimmer\_rel = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A^i - A^{i+1}|}{\frac{1}{N} \sum_{i=1}^N A^i}$$

dove  $A^i$ ,  $i = 1, 2, \dots, N$  sono le ampiezze picco-picco,  $N$  è il numero di impulsi stimati.

- APQ3 (%): misura della variazione dell'ampiezza picco-picco da periodo a periodo con una media lungo 3 periodi successivi. Si calcola come:

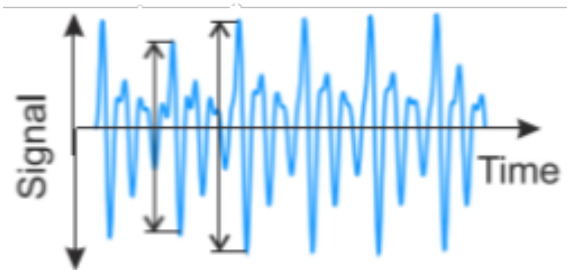
$$APQ3 = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| \frac{A^{i-1} + A^i + A^{i+1}}{3} - A^i \right|}{\frac{1}{N} \sum_{i=1}^N A^i}$$

dove  $A^i$ ,  $i = 1, 2, \dots, N$  sono le ampiezze picco-picco,  $N$  è il numero di impulsi stimati.

- APQ5 (%): misura della variazione dell'ampiezza picco-picco da periodo a periodo con una media lungo 5 periodi successivi. Si calcola come:

$$APQ5 = \frac{\frac{1}{N-4} \sum_{i=3}^{N-2} |A^{i-2} + A^{i-1} + A^i + A^{i+1} + A^{i+2} - 5A^i|}{\frac{1}{N} \sum_{i=1}^N A^i}$$

dove  $A^i$ ,  $i = 1, 2, \dots, N$  sono le ampiezze picco-picco,  $N$  è il numero di impulsi stimati.

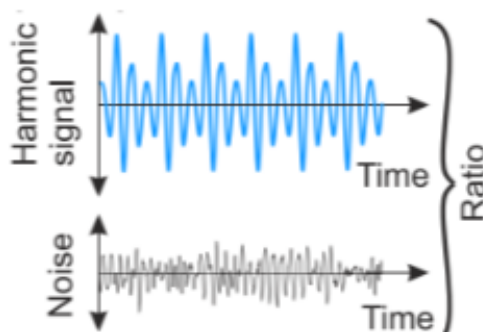


**Figura 3.3:** Misura di shimmer in un segnale vocale [8].

- Spettrali: le misure basate sullo spettro sono in grado di fornire misure del grado di perturbazione del segnale[47]. Esempi di misure spettrali possono essere:
  - *Spectral center of gravity*: Il centro di gravità spettrale è una misura di quanto siano alte in media le frequenze in uno spettro [39];
  - *Standard deviation*: la deviazione standard è una misura di quanto le frequenze in uno spettro possano deviare dal centro di gravità [39];
  - *Skewness*: è una misura della simmetria dello spettro intorno al centro di gravità [39];
  - *Kurtosis*: è una misura di quanto la forma dello spettro attorno al centro di gravità si discosti da una forma gaussiana [39].
- Cepstrali: Il cepstrum (trasformata di Fourier dello spettro di potenza del segnale vocale) è una misura normalizzata dell'ampiezza di picco cepstrale, che confronta il livello di organizzazione armonica del discorso con il rumore di fondo [47]. Parametro significativo a tal fine è il *Cepstral peak prominence (CPP)* che risulta in forte correlazione con la mancanza del respiro sia nelle vocali sostenute che nelle frasi. Altri parametri che rientrano in questa famiglia sono i Mel Frequency Cepstral Coefficients (MFCC): il loro obiettivo

è la rilevazione dei sottili cambiamenti che si verificano nel movimento delle articolazioni del tratto vocale, come la lingua, la mascella e le labbra [48].

- Harmonic to noise ratio (HNR): è definito come la quantità di rumore nel segnale vocale, e modella la voce rauca. Il calcolo di questa misura si basa su un metodo di autocorrelazione che consente di determinare la frequenza e la posizione di ogni ciclo della vibrazione delle corde vocali [49] (Figura 3.4).



**Figura 3.4:** Misura di HNR in un segnale vocale [8].

- Linear predictive coefficient (LPL): possono essere considerati come una codifica percettiva del segnale vocale. La procedura, che mira a replicare i meccanismi percettivi dell'udito, consente di modellare la funzione di trasferimento del tratto vocale a partire dal segnale vocale [50]
- Zero Crossing Rate (ZCR): indica la frequenza a cui il segnale attraversa l'asse del tempo a livello di ampiezza zero, ovvero il numero di volte che il segnale cambia di segno durante un determinato intervallo di tempo. ZCR può fornire informazioni sulla frequenza di transizione tra le diverse vocali e consonanti, sulla presenza di rumore o distorsioni nella voce, ed è ampiamente utilizzato per valutare il tremore nella voce nei pazienti con PD [51].

Nel dataset LUHS sono state estratte 18 tipologie di set di feature, di cui solo 17 utilizzate in questo lavoro di tesi, dal momento che i valori forniti per un particolare subset risultano essere uguali per tutte le registrazioni, in seguito ad un probabile errore in fase di creazione. I dettagli di questi set di feature vengono forniti nella Tabella 3.3. I set di feature da 1 a 13 (tranne 6) sono stati estratti utilizzando *OpenSMILE toolkit* [52], mentre il set di feature 6 è stato estratto utilizzando *Essentia*, una nota libreria C++ sviluppata per scopi di analisi vocale [53]. Il set di feature 15 è stato estratto utilizzando una libreria basata su Java chiamata *MPEG7AudioEnc* [54].

Le features "*jAudio*" (14) sono state estratte tramite una libreria basata su Java, *jAudio* [55]. Il set di features "*YAFEE*" è stato estratto utilizzando il toolbox

*YAFEE* [56]. Infine, un set di features basate sulle misure tempo-frequenza sono state estratte e denominate "*Tsanas*" [46].

**Tabella 3.3:** Tabella riassuntiva set di features del dataset LUHS.

No	Feat. Extr.	Abbreviazione	Toolkit
1	Avec2011	AV1	OpenSMILE toolkit [52]
2	Avec2013	AV2	OpenSMILE toolkit [52]
3	Emo_large	EL	OpenSMILE toolkit [52]
4	Emobase	EM1	OpenSMILE toolkit [52]
5	Emobase2010	EM2	OpenSMILE toolkit [52]
6	Essentia_descriptors	ED	Essentia [53]
7	IS09_emotion	IS1	OpenSMILE toolkit [52]
8	IS10_paraling	IS2	OpenSMILE toolkit [52]
9	IS10_paraling_compat	IS3	OpenSMILE toolkit [52]
10	IS11_speaker_state	IS4	OpenSMILE toolkit [52]
11	IS12_speaker_trait	IS5	OpenSMILE toolkit [52]
12	IS12_speaker_trait_compat	IS6	OpenSMILE toolkit [52]
13	IS13_ComPare	IS7	OpenSMILE toolkit [52]
14	jAudio_features	JA	jAudio [55]
15	MPEG7_descriptors	MP	MPEG7AudioEnc [54]
16	Tsanas	TS	Tsanas [46]
17	YAAFE_features	YA	YAAFE toolbox [56]

Per il dataset AAPP è stato scelto come metodo di estrazione *IS13-ComPare* (dal toolkit di openSMILE)[52], dal momento che tra tutti risulta essere quello comprendente il maggior numero di features.

### 3.3.1 OpenSMILE features

I set di feature 1-12 sono stati ottenuti utilizzando il toolkit *OpenSMILE* (version 2.2 RC 1) [52]. Il nome di ogni set di feature è uguale al nome del file di configurazione (.conf). Dal momento che le varie features estratte sono simili, per semplicità si riporta solo il contenuto di IS13-ComPare.conf (Tabella 3.4).



**Tabella 3.4:** Tabella riassuntiva features ComPare [52].

IS13 ComPare
Ratio of non-zero values, Segment lengths, loudness, Modulated loudness, RMS Energy, ZCR, RASTA auditory bands 1–26, MFCC 1–14, Band energy, Band energy, Spectral flux, Spectral centroid, Spectral entropy, Spec. moments, Spectral slope, Harmonicity, Sharpness (spectral), F0, Prob. of voicing, Jitter, Shimmer, log.HNR

### 3.3.2 Essentia descriptors

Il set di features n.13 è stato ottenuto utilizzando una libreria C++ open source per l'analisi audio: Essentia [53] (versione 2.1 beta 2). Le features in totale sono 1915, di cui 17 globali (descrittori long-term o di alto livello) e  $146 \times 13$  locali (descrittori short-term o di basso livello). Si dividono nelle seguenti famiglie (Tabella 3.5):

- Descrittori low-level;
- Descrittori tonali;
- Descrittori ritmici;

### 3.3.3 jAudio features

Il set di features n. 14 è stato calcolato utilizzando l'applicazione Java jAudio (versione 0.4.5.1) [55]. Si fornisce un elenco dettagliato delle features: 100 stime dell'area (momenth zero), 13 MFCC, 10 LPC, 4 spettrali (centroide, flusso, punto di rolloff, variabilità), 3 strongest frequency, 2 basate sullo spettro (centroide, flusso), smoothness, compactness, root mean square, fraction of low energy windows, relative difference function, e zero crossings basate sul picco spettrale.

**Tabella 3.5:** Tabella riassuntiva delle famiglie di features Essentia [53].

Essentia Descriptors		
LOW-LEVEL		
loudness_ebu128	average_loudness	
dynamic_complexity	silence_rate_20dB	
silence_rate_30dB	silence_rate_60dB	
spectral_rms	spectral_flux	
spectral_centroid	spectral_kurtosis	peactal_spread
spectral_skewness	spectral_rolloff	spectral_decrease
hfc_spectral_strongpeak	zeroocrossingrate	spectral_energy
spectral_energyband_low	spectral_energyband_middle_low	
spectral_energyband_middle_high		
spectral_energyband_high	barkbands_melbands	
melbands128	mfcc_gfcc	
barkbands_crest		
barkbands_flatness_db		
barkbands_kurtosis	barkbands_skewness	
barkbands_spread		
melbands_crest	melbands_flatness_db	
melbands_kurtosis	melbands_skewness	
melbands_spread	erbbands_flatness_db	erbbands_kurtosis
erbbands_skewness	erbbands_spread	dissonance_spectral_entropy
pitch_salience		
spectral_complexity	spectral_contrast_coefs	spectral_contrast_valleys
TONALI		
tuning_frequency	hpcp	
thpcp	hpcp_entropy	hpcp_crest
key_temperley	key_krumhansl	
key_edna	chords_strength	
chords_histogram	chords_changes_rate	
chords_number_rate		
chords_key	chords_scale	
tuning_diatonic	strength	tuning_equal_tempered_deviation
tuning_nontempered	energy_ratio	
RITMICI		
beats_position	beats_count	
bpm_histogram	first_peak_bpm	
bpm_histogram	first_peak_spread	
bpm_histogram	first_peak_weight	
bpm_histogram	second_peak_bpm	
bpm_histogram	second_peak_spread	
bpm_histogram	second_peak_weight	
beats_loudness	band_ratio	
onset_rate	danceability	

### 3.3.4 MPEG7 descriptors

Il set di features n.15 è composto da descrittori basati sullo standard MPEG-7, estratti utilizzando la libreria Java MPEG7AudioEnc [54](versione 0.4 RC 3). Risulta essere composto da:

- 7 descrittori globali: 4 spettrali armonici (centroide, deviazione, variazione, diffusione), 2 centroidi (spettrali, temporali) e log attack time;
- 40 descrittori locali: 36 spettro audio (24 flatness, 10 involuppo, centroide, spread), 2 harmonicity, frequenza fondamentale e potenza.

### 3.3.5 Tsanas features

Il set di features YAFEE (n. 16) contiene varie misure di disfonia ed era inizialmente dedicato specificamente allo screening del PD. Il codice Matlab per calcolare queste funzionalità è pubblicamente disponibile come Voice Analysis Toolbox (versione 1.0) e l'elenco completo delle 339 features è descritto nella tesi di dottorato di *Athanasios Tsanas* [46]. Sono presenti varianti jitter, varianti shimmer, rapporto armonico-rumore, rapporto rumore-armoniche, quoziente glottale, rapporto eccitazione glottale-rumore, rapporto eccitazione corde vocali, log energy, 13 MFCC e loro 1a e 2a derivata, analisi di fluttuazione de-trend, entropia del periodo di pitch.

### 3.3.6 YAAFE features

Il set di features n. 17 è stato ottenuto tramite toolbox per l'estrazione di funzionalità audio: YAAFE [56] (versione 0.65). Le impostazioni predefinite sono rimaste intatte per il seguente elenco di features: 24 loudness, 23 fattore di cresta spettrale per banda, 23 spectral flatness per banda, 13 MFCC, 12 statistiche di forma (4 involuppo, 4 spettrali, 4 temporali), 10 LSF, 10 Octave Band Signal Intensities (OBSI), 9 Octave Band Signal Intensities Ratios (OBSIR), 8 modulazione di ampiezza, 6 spettrale (diminuzione, planarità, flusso, rolloff, pendenza, variazione), 2 LPC, sharpness, spread, energia e zero-crossing rate.

## 3.4 Classificazione PD

Il Dataset LUHS è stato utilizzato in questo lavoro di tesi per la definizione di un algoritmo di classificazione; questo si articola negli step descritti in Figura 3.5:

Tutto l'algoritmo di classificazione è stato applicato in parallelo alle registrazioni acquisite tramite smartphone e tramite microfono.

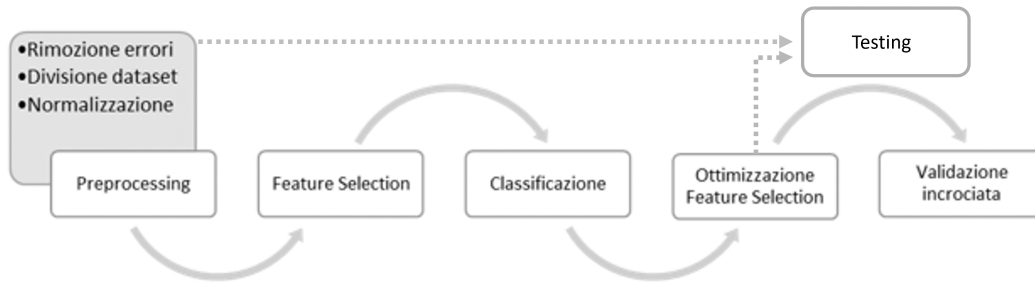


Figura 3.5: Flowchart algoritmo classificazione

### 3.4.1 Preparazione dei dati e dei segnali

Dal momento che nel dataset alcune registrazioni presentano errori di *mancata corrispondenza* tra ID/classe microfono e ID/classe smartphone, queste sono state eliminate, con lo scopo di evitare che l'algoritmo di classificazione possa essere allenato su classi errate.

Di queste registrazioni rimanenti, è stata presa in considerazione solo la prima per ogni paziente.

Il passo successivo è stata la divisione del dataset in training-set e test-set.

- Training-set: è il subset del dataset originale che viene utilizzato per addestrare il modello. È la porzione del dataset che il modello "vede" durante la fase di apprendimento, e viene utilizzata per calibrare i parametri del modello.
- Test-set: è un subset separato del dataset originale che viene utilizzato per valutare le prestazioni del modello. Una volta che il modello è stato addestrato utilizzando il training set, viene valutato sul test set per misurare la sua capacità di generalizzazione e di predizione su nuovi dati. Il test set non viene utilizzato durante la fase di addestramento.
- Validation-set: è un subset aggiuntivo del dataset originale che viene utilizzato per regolare gli iperparametri del modello, parametri che ne controllano la complessità. Il validation set viene utilizzato per scegliere i migliori valori di questi iperparametri, che massimizzano la capacità del modello di generalizzare su nuovi dati.

Per evitare una debole capacità di generalizzazione dell'algoritmo, si è deciso di dividere il dataset randomicamente adottando la seguente ripartizione bilanciata per classe:

- Training-set: 80% del dataset di partenza
- Test-set: 20% del dataset di partenza

- Validation-set: è stata utilizzata una tecnica di tuning dei dati, ovvero la *k-fold crossvalidation (CV)*, utilizzata per valutare i modelli di machine learning in presenza di un campione di dati limitato. La procedura ha un singolo parametro chiamato  $k$  ( $k=10$ ) che si riferisce al numero di gruppi in cui deve essere suddiviso il training-set. Il funzionamento della *k-fold cross-validation* può essere descritto come segue:
  - Si divide il dataset in  $k$  parti (o "fold");
  - Si addestra il modello su  $k-1$  fold del dataset (il training set) e si valuta sul fold rimanente (il validation set).
  - Si ripete questo processo  $k$  volte, utilizzando ogni volta un fold differente come validation set.
  - Si calcola la media delle prestazioni del modello su tutti i validation set. Questo valore medio viene considerato come una stima dell'accuratezza del modello sul dataset completo.

I due subset, training e test-set, sono stati scelti facendo attenzione a garantire che le registrazioni contenute nel training non fossero anche nel test. La *k-fold CV* è stata applicata sul training set. Passo successivo è stata la *normalizzazione* di training-set e test-set con lo scopo di evitare che la distanza speaker-microfono influisse sul modello di classificazione. E' stata utilizzata la *z score standardization*, che riscalda i valori in modo tale da ottenere una media di 0 e una deviazione standard di 1.

$$z = \frac{x - m}{\sigma} \quad (3.1)$$

Nella formula 3.1,  $x$  corrisponde alle istanze contenute nel dataset,  $m$  è la media dei valori del training-set e  $\sigma$  è la deviazione standard dei valori del training-set.

### 3.4.2 Feature Selection

In generale, un algoritmo di feature selection fornisce un subset costituito solo dalle features rilevanti e non correlate, permettendo la riduzione delle dimensioni del dataset di partenza [35].

Sono stati utilizzate tre metodi di feature selection:

- ANOVA: è l'acronimo di *analisi della varianza* ed è un test di ipotesi statistica parametrica per determinare se le medie di due o più campioni di dati (spesso tre o più) provengono dalla stessa distribuzione o meno. I risultati di questo test indicano le feature indipendenti dalla variabile target e che, quindi, possono essere rimosse dal set di dati. Questa tipologia di feature selection dunque

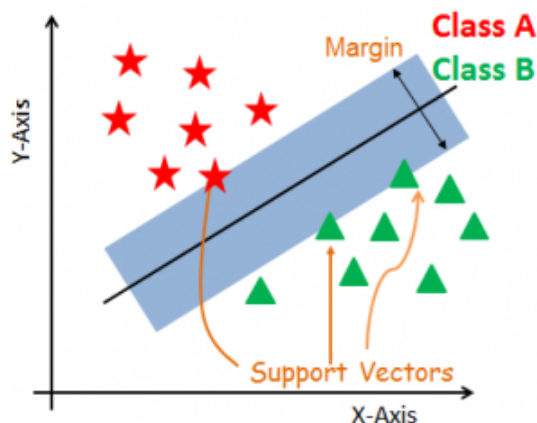
restituisce le features in base ai k punteggi più alti. Si è deciso di settare il valore di k a 10.

- **BORUTA**: è un metodo di selezione di tutte le features rilevanti, al contrario di altri metodi *minimi ottimali*; ciò significa che cerca di trovare tutte le features che contengono informazioni utilizzabili per la previsione, piuttosto che trovare un sottoinsieme compatto di features su cui il classificatore ha un errore minimo [57]. L'algoritmo di Boruta viene identificato come metodo di feature selection *wrapper* costruito attorno all'algoritmo di classificazione *Random Forest (RF)* [58]. In un metodo wrapper il modello viene quindi utilizzato come una scatola nera che restituisce una classifica delle features sulla base del loro potere di classificazione.
- **Correlation-Based Feature Selection**: questa procedura ha lo scopo di selezionare le features più significative (ovvero quelle con un'elevata correlazione features-classe) e non ridondanti (cioè con bassa correlazione tra le features stesse). Innanzitutto, è stato valutato il coefficiente di Pearson ( $r$ ) tra feature e classe ( $r_{f0}$ ), indagando il suo valore assoluto per ogni feature e conservando solo il massimo di quelli significativi (cioè,  $r > 0.3$ , P - value  $< 0.05$ ). È stata successivamente calcolata la correlazione intra-feature ( $r_{ff}$ ), e per coppie di features che mostrano un'inter-correlazione superiore all'intra-correlazione (cioè,  $r_{ff} > r_{f0}$ ), la feature meno correlata alla classe viene rimossa [59].

### 3.4.3 Classificazione

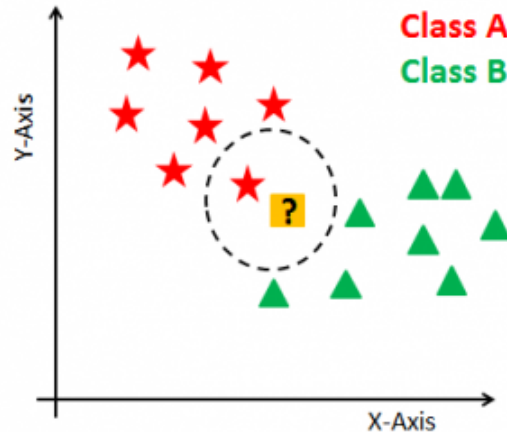
Sono stati implementati quattro modelli di classificazione: *SVM*, *KNN*, *GNB* e *GB*, e con ognuno di questi è stata eseguita una 10-fold cross-validation utilizzando i subset di features derivati dallo step di Feature Selection. Andando nel dettaglio del funzionamento dei classificatori:

- **SVM**: modello di classificazione il cui obiettivo è quello di trovare la retta di separazione delle classi che massimizza il margine tra le classi stesse, dove con margine si intende la distanza minima dalla retta ai punti delle due classi. Questo obiettivo viene raggiunto utilizzando una parte minimale del dataset di allenamento, i cosiddetti vettori di supporto (da cui il nome della famiglia di modelli). I vettori di supporto sono i valori di una classe più vicini alla retta di separazione, quelli che più si avvicinano all'altra classe. [60] I parametri di input del classificatore sono:
  - **Kernel**: definisce se si vuole adottare una separazione lineare o non lineare. Questa scelta dipende dalla complessità della disposizione dei valori nel piano [60].



**Figura 3.6:** Funzionamento di base del Support Vector Machine classifier [61].

- Gamma: definisce fino a che punto arriva l’influenza di un singolo esempio di allenamento; in altre parole, con valori di gamma bassi, i punti lontani dalla linea di separazione plausibile vengono considerati nel calcolo della linea di separazione, mentre per valori alti, solo i punti più vicini alla linea plausibile vengono considerati nel calcolo [60].
- Regularization (C): tramite questa variabile si può controllare la penalità nel caso di errore di classificazione. Valori grandi di C corrispondono ovviamente a penalità grandi e quindi si costringe il modello ad avvicinarsi il più possibile al training set; invece si è meno stringenti a riguardo di errori di classificazione con valori più piccoli di C [60].
- *KNN*: è un classificatore di apprendimento supervisionato non parametrico che utilizza la prossimità per effettuare classificazioni o previsioni sul raggruppamento di un singolo punto. Per i problemi di classificazione, viene assegnata una etichetta di classe sulla base di un voto di maggioranza, ovvero viene utilizzata l’etichetta rappresentata più frequentemente attorno a un punto. Il valore  $k$  nell’algoritmo  $kNN$  definisce quanti vicini verranno controllati per determinare la classificazione di un punto specifico. Ad esempio, se  $k=1$ , la classe verrà assegnata come la stessa classe del suo singolo più vicino, come mostrato nella figura 3.7. Definire  $k$  può essere un atto di bilanciamento poiché valori diversi possono portare a overfitting o underfitting [62].
- *GNB*: il Gaussian Naïve Bayes è un algoritmo probabilistico di apprendimento automatico utilizzato per molte funzioni di classificazione e si basa sul teorema di Bayes; risulta essere una estensione del Naïve Bayes. Il nome "Naïve" viene utilizzato perché l’algoritmo incorpora features nel suo modello che sono indipendenti l’una dall’altra. Eventuali modifiche al valore di una feature



**Figura 3.7:** Funzionamento di base del K-Nearest Neighbors classifier, con un valore di  $k=1$

dunque non influiscono direttamente sul valore di qualsiasi altra dell'algoritmo. Il vantaggio principale è che si tratta di un algoritmo semplice ma potente [63].

- *GB*: Il Gradient Boosting è un classificatore che utilizza un tipo speciale di tecnica di Ensemble Learning, il *Boosting*, che funziona combinando predittori con scarsa accuracy con modelli con una alte accuracy. I due metodi di potenziamento più popolari sono: Adaptive Boosting e gradient boosting [64]. In GB, ogni predittore cerca di migliorare attraverso il suo predecessore riducendo gli errori, per cui l'idea affascinante alla base del GB è che invece di adattare un predittore ai dati ad ogni iterazione, in realtà adatta un nuovo predittore agli errori residui commessi dal predittore precedente [64]. I parametri più importanti da settare sono i seguenti [64]:
  - Criterio: la funzione di loss utilizzata per trovare la features e la soglia ottimale per suddividere i dati; il ruolo della funzione di loss è la stima della capacità del modello di fare previsioni con i dati forniti.
  - Learning-rate: scala il contributo di ciascun albero.
  - Max-depth: la profondità massima di ciascun albero.
  - N-estimators: il numero di alberi da costruire.

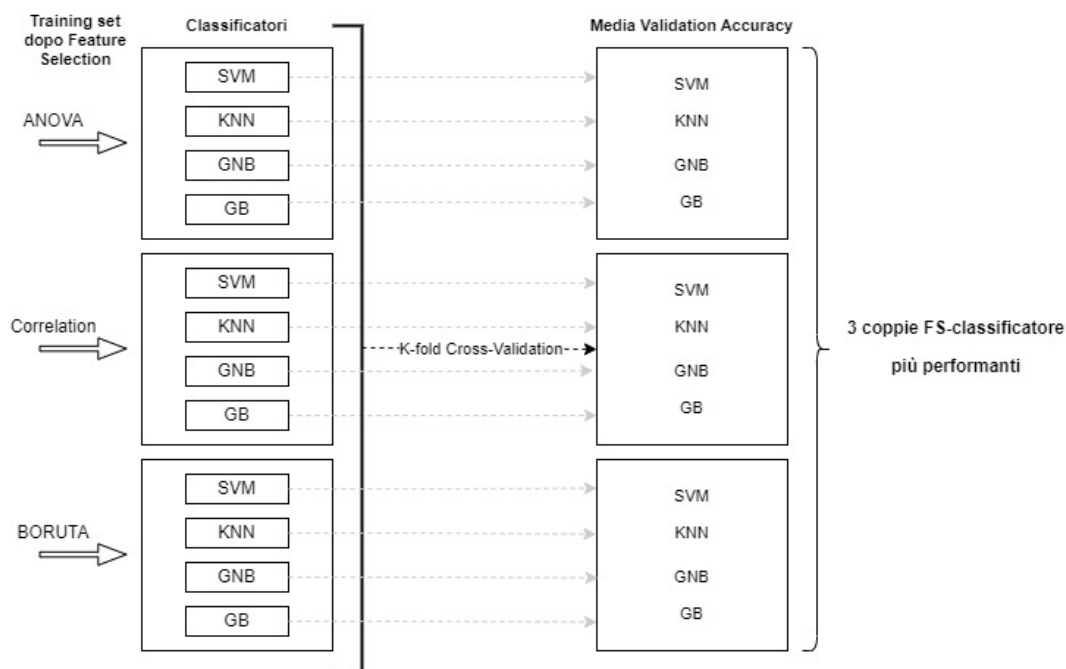
Per ognuno di questi classificatori sono stati utilizzati come parametri di input quelli di default.

Poiché il set di dati è bilanciato (ovvero la cardinalità è la stessa in ogni classe), come metrica di valutazione delle prestazioni si è considerata l'*accuratezza*, definita



come il rapporto tra il numero delle predizioni corrette e il numero totale di predizioni.

Per ogni classificatore sono stati salvati tutti i valori di accuratezza restituiti dalla k-fold cross-validation e di questi valori è stata calcolata la media. Sulla base del valore più alto di media si sceglie con quale classificatore continuare, ovvero quale classificatore performa meglio per le 3 tipologie di Feature Selection considerate. Dunque otterremo le 3 coppie Feature Selection-classificatore più performanti, di cui sceglieremo ancora una volta quella con accuratezza sul validation più alta per proseguire con una ottimizzazione della Feature Selection finale. Per facilitare la comprensione di tale processo, viene fornito uno schema esemplificativo in Figura 3.8.



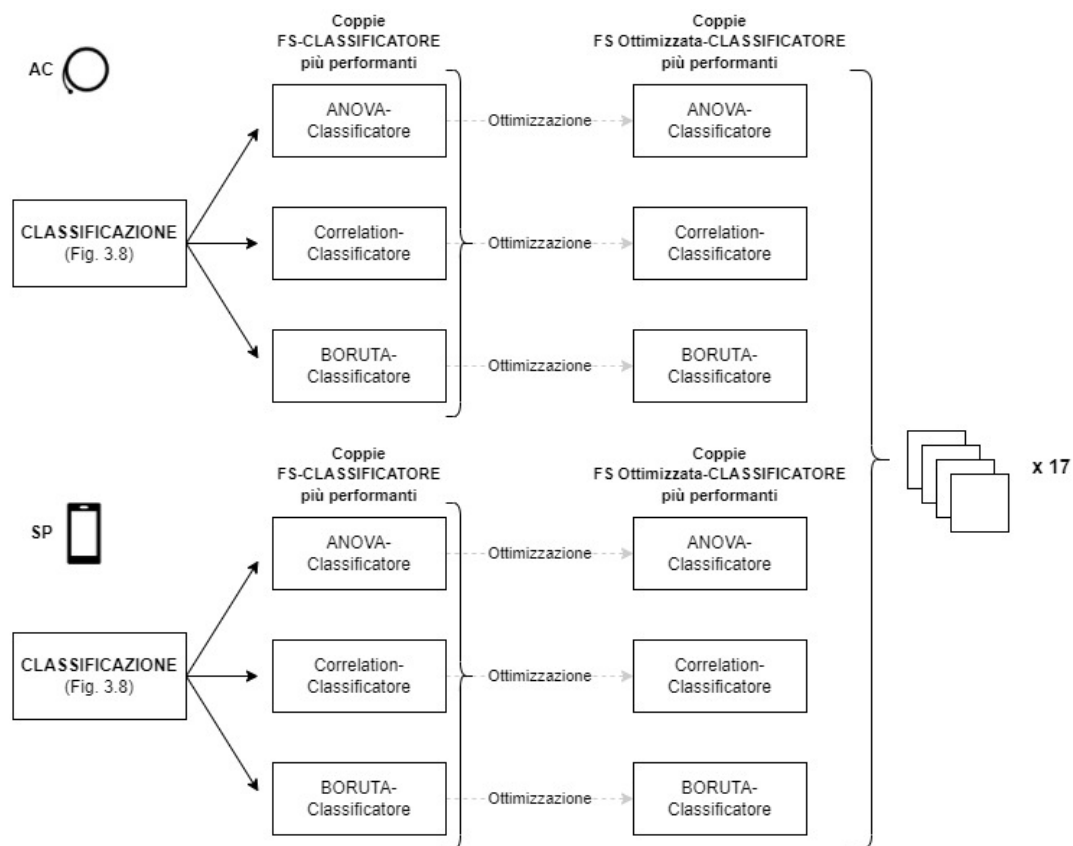
**Figura 3.8:** Flowchart nel dettaglio dell’algoritmo di classificazione.

Passando alla ottimizzazione della Feature Selection, sono state attuate le seguenti strategie:

- ANOVA: è stato fatto variare da 5 a 40 (a passi di 5) il numero di features con i k punteggi più alti. Variando il numero di k si ha dunque la possibilità di includere (aumentando k) features rappresentative e informative oppure escludere features che potrebbero non essere informative per un particolare set di dati.

- Correlation-based approach: la soglia della correlazione di Pearson  $r$  è stata aumentata da 0.3 a 0.4 a passi di 0.01. In questo modo la soglia diventa più stringente e permette l'inclusione solo delle features maggiormente correlate con la classe.
- BORUTA: è stato modificato (abbassato) il parametro percentuale che regola il numero di falsi positivi selezionati, consentendo all'algorithm di includere features che potrebbero essere informative.

Ultimo step dell'ottimizzazione è stata la scelta, ancora una volta attraverso il valore della validation accuracy più alto, del parametro migliore modificato della feature selection.

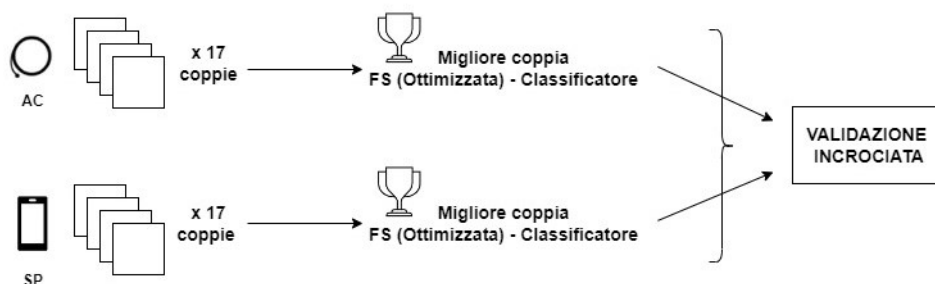


**Figura 3.9:** Flowchart nel dettaglio dell'ottimizzazione della Feature Selection, a partire dai risultati ottenuti dalla Classificazione.

Arrivati a questo punto finale dell'algorithm di classificazione, ciò che si è ottenuto è, per ognuno dei 17 set di dati (sia AC che SP), una coppia Feature Selection (ottimizzata)- classificatore migliore, che ha raggiunto risultati più alti sul validation set. Ciò viene mostrato nella Figura 3.9.

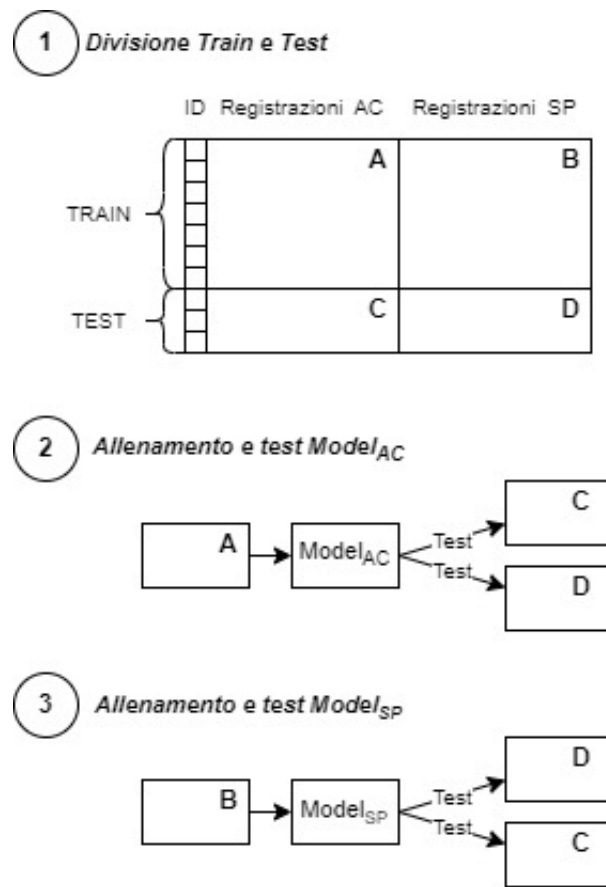
### 3.4.4 Validazione incrociata

Lo scopo di questo lavoro di tesi è quello di validare le registrazioni vocali acquisite con smartphone, ed è proprio con questo fine che si è svolta una fase di validazione incrociata tra la migliore coppia FS-classificatore per il microfono professionale, ( $Model_{AC}$ ) e la migliore coppia FS-classificatore per lo smartphone, ( $Model_{SP}$ ). Tale analisi è stata condotta testando il modello allenato sui dati registrati tramite microfono su quelli registrati tramite smartphone (e viceversa), con lo scopo di valutare l'influenza delle condizioni di registrazioni sul modello costruito.



**Figura 3.10:** Descrizione utile alla comprensione dei dati utilizzati per la validazione incrociata.

Al fine di evitare bias nei risultati, il dataset è stato preliminarmente suddiviso in train e test: gli id dei soggetti sono stati randomicamente assegnati al gruppo di train o a quello di test, quindi le registrazioni corrispondenti sono state utilizzate per l'allenamento e per il test del modello. Uno schema della procedura eseguita è mostrato in Figura 3.11.



**Figura 3.11:** Schema della procedura di validazione incrociata.

Vale la pena di notare che, per fare in modo che la procedura di selezione randomica dei soggetti restituisse sempre lo stesso set di id, è stato originariamente settato un seed pari a 3.

Il risultato di tale analisi consente di valutare quanto l'applicazione di un algoritmo allenato su dati raccolti con una certa modalità subisca una variazione, in termini di accuratezza, quando applicato su campioni registrati con una strumentazione differente.

### 3.5 Analisi statistica

Con lo scopo di identificare quali feature estratte dai segnali vocali siano maggiormente invarianti rispetto al sistema di registrazione utilizzato, si è portato avanti un test statistico, ovvero il *Wilcoxon signed-rank test* [65]. Questo test statistico è uno strumento non parametrico che confronta due gruppi appaiati ovvero, nel caso di questo lavoro di tesi, le features derivanti dalle registrazioni SP e AC.

Le ipotesi del test di Wilcoxon signed-rank sono le seguenti:

- Ipotesi nulla ( $H_0$ ): non ci sono differenze significative tra i due campioni appaiati;
- Ipotesi alternativa ( $H_1$ ): ci sono differenze significative tra i due campioni appaiati.

Più nel dettaglio, il test confronta le differenze tra le coppie di osservazioni nei due campioni appaiati:

1. Viene calcolata la differenza tra le coppie di osservazioni nei due campioni appaiati.
2. Le differenze sono classificate in ordine di valore assoluto, ovvero in ordine crescente di distanza dal valore zero.
3. Viene assegnata una classifica a ciascuna differenza in base al loro ordine di grandezza.
4. Si calcola il valore di test, che rappresenta la somma delle classifiche delle differenze positive o negative, a seconda del segno della differenza. Se il valore di test è sufficientemente alto, viene rifiutata l'ipotesi nulla che i due campioni appaiati siano stati estratti dalla stessa popolazione. In altre parole, ci sono differenze significative tra i due campioni.

Per determinare la significatività del valore di test, viene utilizzata una tabella di valori critici, che dipende dalla dimensione del campione e dal livello di significatività scelto.

Il *p-value* nel test è un parametro fondamentale per interpretare i risultati e indica la probabilità di ottenere una differenza tra i due campioni appaiati che sia uguale o più grande di quella osservata, se non ci sono differenze reali tra i campioni.

Il test restituisce una percentuale di confidenza (p-value) che dovrà essere interpretato nel seguente modo:

- Se il p-value è inferiore al livello di significatività alpha (in genere si considera  $\alpha=0.05$ ), allora viene rifiutata l'ipotesi nulla e si conclude che ci sono differenze significative tra i due campioni appaiati;
- Se il p-value è maggiore del livello di significatività scelto, allora non si può concludere che ci siano differenze significative tra i due campioni appaiati. In altre parole, quando il p-value è maggiore di alpha, non si rifiuta l'ipotesi nulla che i due campioni appaiati siano stati estratti dalla stessa popolazione e non si può accettare l'ipotesi alternativa che i due campioni differiscano in modo significativo tra loro.

Si è partiti dal Dataset LUHS: su ogni set di dati (dall'1 al 17) è stato eseguito il Wilcoxon signed-rank test che ha fornito per ogni features un valore di p-value. Questi valori sono stati ordinati sia in ordine di p-value crescente che decrescente, in modo tale da selezionare le prime 20 features statisticamente differenti tra AC e SP ( $p\text{-value} < 0.05$ ) e le prime 20 features che non presentano differenza statisticamente significativa ( $p\text{-value} > 0.05$ ), rispettivamente.

Da qui seguono analisi approfondite su quali famiglie di features riportano valori statisticamente non differenti tra le due modalità di acquisizione AC e SP e che, conseguentemente, potrebbero essere invarianti rispetto alla tecnica di registrazione scelta.

Dopo aver identificato, per ciascun set di dati del dataset LUHS (Tabella 3.3), le feature che presentano maggiori e minori differenze tra le due condizioni di registrazioni, è stata effettuata un'analisi cross-toolbox: partendo dal presupposto che i set di dati LUHS, seppur con varie differenze, valutano simili set di feature, è stato effettuato un confronto sui risultati riportati su ciascun set per valutare la robustezza dei risultati. Vale la pena notare che, date le differenze tra i vari toolbox, l'analisi è stata semplificata effettuando il confronto tra "famiglie di feature" ottenute raggruppando descrittori appartenenti ad uno stesso ambito, come descritto di seguito:

- F0 related: raggruppa tutti i parametri statistici relativi alla frequenza fondamentale.
- Jitter: raggruppa tutte le varianti di Jitter.
- Shimmer: raggruppa tutte le varianti di Shimmer.
- Spectral: raggruppa tutte le feature relative all'analisi dello spettro del segnale.
- Cepstral: raggruppa tutte le feature relative all'analisi del cepstrum.
- LPC: raggruppa tutte le feature LPL.
- Noise: raggruppa tutte le feature relative all'analisi del rumore.
- ZCR: raggruppa tutti i parametri statistici relativi allo zero crossing rate.
- Energy in specific bands: raggruppa tutte le feature associate all'energia in bande specifiche dello spettro (vari toolbox identificano regioni significative dello spettro del segnale acustico ed estraono misure dell'energia contenuta nelle suddette bande).
- Probability of voice: raggruppa tutte le feature relative all'alternanza tra segmenti sordi e sonori.

- Area of moments: raggruppa le misure della distribuzione di energia del segnale acustico.
- Loudness: raggruppa tutte le misure dell'intensità sonora del segnale acustico.
- Wavelet based features: raggruppa tutte le misure calcolate a partire dalla trasformata wavelet del segnale acustico.
- Envelope Descriptors: raggruppa tutti i descrittori dell'involuppo del segnale acustico.

Seppur con evidenti limitazioni, tale analisi ha consentito di valutare quanto le singole feature vengano influenzate dalla tecnica di acquisizione, in maniera indipendente dall'algoritmo utilizzato per il loro calcolo (toolbox diversi presentano spesso algoritmi diversi per la valutazione della stessa feature).

A partire dai risultati ottenuti sul dataset LUHS, il dataset AAPP è stato utilizzato per una fase di validazione. Il set di feature ComPare (13) è stato estratto a partire dai segnali audio contenuti in AAPP. Come specificato in 3.3.2, tale set è stato scelto in quanto più completo. Di tale subset (13) sono state considerate le 20 features che presentano maggiori differenze, quindi p-value basso: ciò è utile a comprendere quale famiglia di features risulta essere maggiormente affetta dalla modalità di registrazione, con lo scopo di ideare una metodologia che possa minimizzarne le differenze.





# Capitolo 4

## Risultati

### 4.1 Classificazione PD

Vengono riportati nelle tabelle, per ogni subset di dati (da 1 a 17), il valore massimo delle medie di accuracy del validation-set che la k-fold crossvalidation restituisce, sia per registrazioni AC (Tabella 4.1) che SP (Tabella 4.2), e il relativo classificatore.

**Tabella 4.1:** Valori di Validation Accuracy massima per ogni subset al variare della Feature Selection utilizzata, per registrazioni AC. Si evidenziano i valori maggiori di Validation Accuracy per ogni subset.

No		ANOVA	BORUTA	CORRELATION
1	max_val_acc classificatore	82.32% SVM	82.50% GB	<b>86.07%</b> <b>SVM</b>
2	max_val_acc classificatore	<b>79.64%</b> <b>SVM</b>	74.82% KNN	71.07% SVM
3	max_val_acc classificatore	78.39% GNB	74.82% KNN	<b>80.89%</b> <b>SVM</b>
4	max_val_acc classificatore	79.82% SVM	<b>83.75%</b> <b>KNN</b>	82.50% KNN
5	max_val_acc classificatore	<b>75.89%</b> <b>SVM</b>	74.64% SVM	65.71% KNN
6	max_val_acc classificatore	80.89% KNN	<b>87.32%</b> <b>GB</b>	83.75% GNB
7	max_val_acc classificatore	68.57% SVM	<b>74.82%</b> <b>KNN</b>	67.14% KNN
8	max_val_acc classificatore	72.32% SVM	<b>74.82%</b> <b>KNN</b>	72.32% SVM
9	max_val_acc classificatore	74.82% GB	<b>82.14%</b> <b>GB</b>	73.21% GB
10	max_val_acc classificatore	81.25% SVM	78.57% SVM	<b>83.57%</b> <b>GNB</b>
11	max_val_acc classificatore	84.64% SVM	84.46% GB	<b>86.07%</b> <b>SVM</b>
12	max_val_acc classificatore	76.07% GB	78.39% GB	<b>79.64%</b> <b>SVM</b>
13	max_val_acc classificatore	76.07% KNN	77.50% SVM	<b>81.07%</b> <b>SVM</b>
14	max_val_acc classificatore	<b>82.50%</b> <b>SVM</b>	81.07% GB	78.39% SVM
15	max_val_acc classificatore	77.32% GB	78.57% SVM	<b>78.58%</b> <b>KNN</b>
16	max_val_acc classificatore	<b>78.21%</b> <b>SVM</b>	77.14% GB	70.89% GNB
17	max_val_acc classificatore	79.82% SVM	<b>83.75%</b> <b>KNN</b>	80.71% GB

**Tabella 4.2:** Valori di Validation Accuracy massima per ogni subset al variare della Feature Selection utilizzata, per registrazioni SP. Si evidenziano i valori maggiori di Validation Accuracy per ogni subset.

No		ANOVA	BORUTA	CORRELATION
1	max_val_acc	<b>79.64%</b>	74.64%	73.39%
	Classificatore	<b>KNN</b>	SVM	SVM
2	max_val_acc	76.07%	77.32%	<b>80.89%</b>
	Classificatore	GB	GB	<b>SVM</b>
3	max_val_acc	76.07%	<b>78.39%</b>	75.89%
	Classificatore	GNB	<b>KNN</b>	KNN
4	max_val_acc	74.11%	<b>78.75%</b>	66.96%
	Classificatore	SVM	<b>GB</b>	GNB
5	max_val_acc	74.82%	<b>80.71%</b>	77.32%
	Classificatore	KNN	<b>SVM</b>	KNN
6	max_val_acc	74.46%	<b>78.75%</b>	78.39%
	Classificatore	GNB	<b>KNN</b>	SVM
7	max_val_acc	70.71%	<b>81.07%</b>	68.75%
	Classificatore	KNN	<b>GB</b>	GNB
8	max_val_acc	<b>78.39%</b>	77.32%	74.82%
	Classificatore	<b>SVM</b>	KNN	SVM
9	max_val_acc	77.32%	<b>78.57%</b>	74.46%
	Classificatore	SVM	<b>SVM</b>	SVM
10	max_val_acc	77.32%	79.82%	<b>83.57%</b>
	Classificatore	SVM	SVM	<b>GNB</b>
11	max_val_acc	82.32%	<b>88.57%</b>	82.32%
	Classificatore	KNN	<b>SVM</b>	SVM
12	max_val_acc	83.57%	81.07%	<b>84.82%</b>
	Classificatore	KNN	SVM	<b>SVM</b>
13	max_val_acc	86.07%	<b>86.25%</b>	85.00%
	Classificatore	SVM	<b>KNN</b>	SVM
14	max_val_acc	70.89%	<b>81.07%</b>	71.07%
	Classificatore	SVM	<b>KNN</b>	GNB
15	max_val_acc	70.54%	<b>75.89%</b>	73.39%
	Classificatore	GNB	<b>GB</b>	SVM
16	max_val_acc	65.54%	<b>79.46%</b>	71.25%
	Classificatore	GNB	<b>SVM</b>	GNB
17	max_val_acc	<b>78.57%</b>	68.75%	76.25%
	Classificatore	<b>SVM</b>	GNB	GNB

Per ogni subset si prosegue con la migliore coppia Feature Selection-classificatore, nonchè la coppia con il valore di validation accuracy più alto, prima di procedere alla *ottimizzazione della Feature Selection*. Vengono riportati i valori di validation accuracy relativi all'ottimizzazione della Feature Selection in 3 tabelle distinte, una per ogni tipologia di Feature Selection: ciò viene effettuato parallelamente sia per AC che per SP.

**Tabella 4.3:** Valori di validation accuracy (AC) al variare del numero di features da selezionare dell'algoritmo *ANOVA*. Si riporta il numero del subset considerato (No) e il classificatore con cui è stato raggiunto tale risultato. Si evidenziano i valori di Validation accuracy migliori.

		Numero Features						
No	Classificatore	5	10	15	20	25	30	35
2	SVM	75.89%	79.82%	78.04%	81.07%	<b>83.57%</b>	82.14%	81.07%
5	SVM	73.57%	74.82%	70.89%	74.64%	<b>74.83%</b>	69.46%	73.57%
14	SVM	78.75%	83.39%	78.21%	<b>83.57%</b>	75.89%	81.07%	80.89%
16	SVM	64.64%	74.64%	78.39%	<b>80.89%</b>	74.82%	69.82%	68.04%

**Tabella 4.4:** Valori di validation accuracy (AC) al variare del valore della soglia della correlazione di Pearson r dell'algoritmo *Correlation-based approach*. Si riporta il numero del subset considerato (No) e il classificatore con cui è stato raggiunto tale risultato. Si evidenziano i valori di Validation accuracy migliori.

		Soglia r							
No	Classificatore	0.3	0.31	0.32	0.33	0.34	0.35	0.36	0.37
1	SVM	86.07%	84.82%	<b>86.25%</b>	83.57%	82.32%	71.07%	-	-
3	SVM	<b>81.25%</b>	80.00%	77.50%	73.39%	-	-	-	-
10	GNB	81.07%	<b>83.57%</b>	75.54%	76.07%	71.07%	74.64%	73.57%	74.82%
11	SVM	<b>88.57%</b>	82.14%	82.14%	77.50%	80.89%	73.57%	-	-
12	SVM	<b>80.71%</b>	78.39%	78.57%	77.14%	76.25%	78.39%	-	-
13	SVM	81.07%	<b>81.96%</b>	78.21%	77.50%	69.82%	73.39%	-	-
15	KNN	78.57%	<b>79.64%</b>	78.21%	65.71%	65.89%	67.32%	-	-

**Tabella 4.5:** Valori di validation accuracy (AC) al variare del valore del parametro percentuale dell’algoritmo *BORUTA*. Si riporta il numero del subset considerato (No) e il classificatore con cui è stato raggiunto tale risultato. Si evidenziano i valori di Validation accuracy migliori.

		Perc.										
No	Class.	90	91	92	93	94	95	96	97	98	99	100
4	KNN	78.75	80.89	81.07	82.14	83.75	82.14	83.75	<b>85.00</b>	80.89	83.57	83.57
6	GB	73.39	71.07	76.07	77.14	78.39	76.25	81.07	79.46	77.14	<b>83.39</b>	81.25
7	KNN	68.39	69.64	74.64	72.14	69.46	71.96	73.57	73.21	73.39	<b>74.83</b>	74.82
8	KNN	73.39	70.54	73.39	<b>78.39</b>	78.24	71.79	74.82	73.39	69.46	65.89	70.71
9	GB	78.39	79.82	69.64	75.89	75.89	70.89	73.39	74.64	75.00	<b>83.39</b>	79.64
17	KNN	82.32	84.82	83.75	78.57	83.57	82.32	85.00	86.07	83.21	<b>86.15</b>	81.96

**Tabella 4.6:** Valori di validation accuracy (SP) al variare del numero di features da selezionare dell’algoritmo *ANOVA*. Si riporta il numero del subset considerato (No) e il classificatore con cui è stato raggiunto tale risultato. Si evidenziano i valori di Validation accuracy migliori.

		Numero Features						
No	Classificatore	5	10	15	20	25	30	35
1	KNN	65.89%	81.07%	76.96%	79.64%	80.00%	80.89%	<b>81.08%</b>
8	SVM	74.64%	<b>77.32%</b>	74.64%	73.57%	75.71%	70.89%	71.07%
17	SVM	73.57%	<b>82.14%</b>	73.39%	75.89%	77.50%	78.57%	79.64%

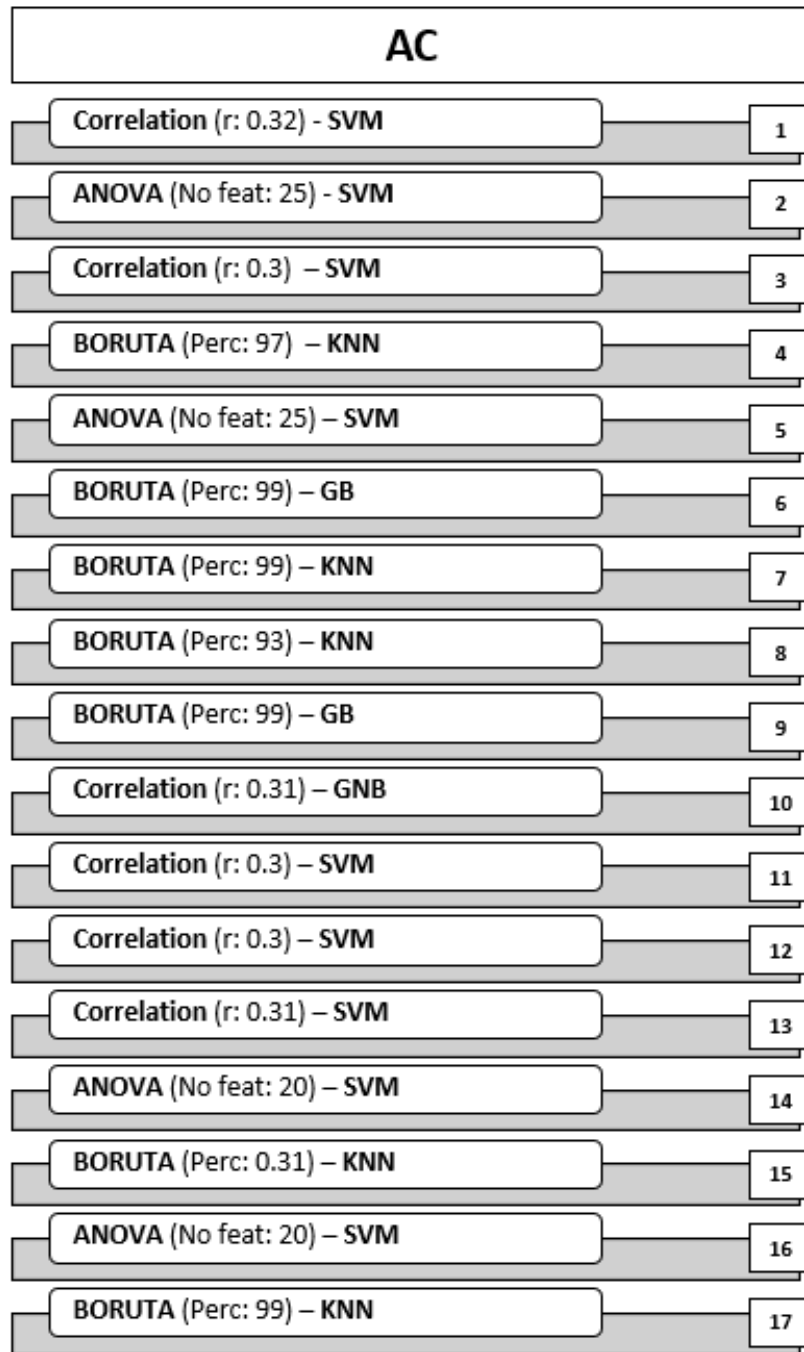
**Tabella 4.7:** Valori di validation accuracy (SP) al variare del valore della soglia della correlazione di Pearson  $r$  dell'algoritmo *Correlation-based approach*. Si riporta il numero del subset considerato (No) e il classificatore con cui è stato raggiunto tale risultato. Si evidenziano i valori di Validation accuracy migliori.

		Soglia $r$						
No	Classificatore	0.3	0.31	0.32	0.33	0.34	0.35	0.36
2	SVM	<b>80.89%</b>	75.89%	78.57%	75.71%	73.39%	70.71%	-
10	GNB	<b>84.82%</b>	82.32%	78.21%	79.64%	77.14%	73.57%	71.96%
12	SVM	<b>84.82%</b>	83.57%	82.32%	82.14%	79.64%	83.21%	82.14%

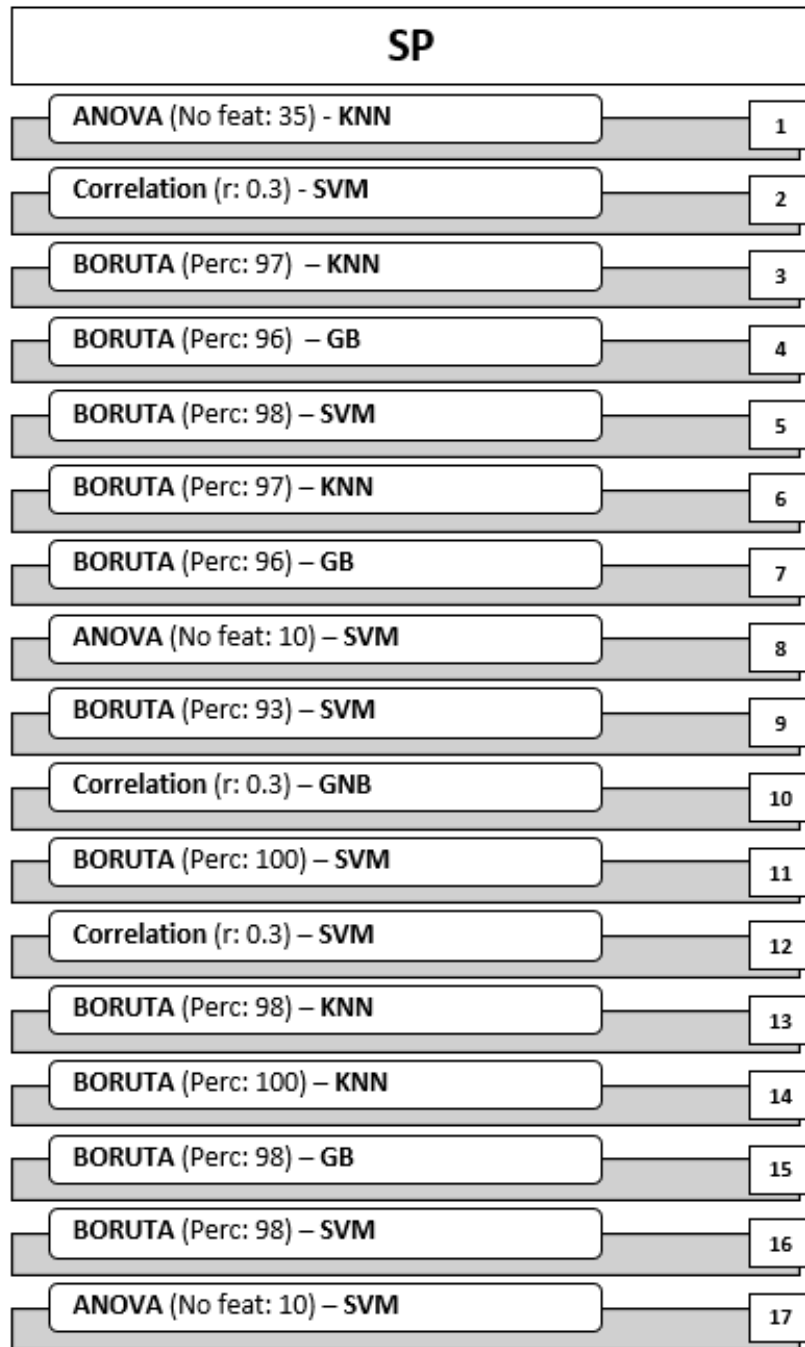
**Tabella 4.8:** Valori di validation accuracy (SP) al variare del valore del parametro percentuale dell'algoritmo *BORUTA*. Si riporta il numero del subset considerato (No) e il classificatore con cui è stato raggiunto tale risultato. Si evidenziano i valori di Validation accuracy migliori.

		Perc.										
No	Class.	90	91	92	93	94	95	96	97	98	99	100
3	KNN	74.64	80.89	74.82	75.89	77.14	77.14	76.07	<b>81.07</b>	77.14	74.82	80.71
4	GB	79.64	79.64	75.89	70.89	72.14	74.82	<b>79.82</b>	76.79	71.96	79.46	74.82
5	SVM	81.25	81.25	83.39	80.89	81.07	82.32	82.14	79.82	<b>84.82</b>	81.07	78.75
6	KNN	77.14	76.07	77.50	74.64	75.89	77.14	73.57	<b>77.51</b>	70.71	74.82	77.32
7	GB	77.32	74.64	76.07	79.82	78.57	84.64	<b>84.82</b>	78.39	73.57	70.89	74.82
9	SVM	81.96	81.07	84.82	<b>84.84</b>	80.89	82.14	80.89	78.39	81.07	80.89	79.82
11	SVM	83.21	83.57	86.25	86.25	86.07	84.82	82.14	82.50	86.25	83.39	<b>88.75</b>
13	KNN	84.82	83.57	83.39	83.57	79.82	86.25	84.64	83.57	<b>88.39</b>	85.00	78.57
14	KNN	70.89	67.14	67.32	67.14	66.07	64.64	64.11	63.39	66.79	77.50	<b>78.57</b>
15	GB	72.14	78.57	72.32	75.54	76.96	78.39	70.89	78.39	<b>79.82</b>	73.39	76.25
16	SVM	74.82	75.71	72.32	72.32	75.00	76.96	77.50	79.82	<b>81.25</b>	79.64	80.00

Con il fine di decretare la coppia migliore *Feature Selection (ottimizzata)-Classificatore* si scelgono per ogni Feature Selection i parametri migliori risultanti dalla fase di ottimizzazione. Si mostrano nelle Figure 4.1 e 4.2 le coppie migliori, sia per AC che per SP, con relativi parametri ottimizzati.



**Figura 4.1:** Migliore coppia Feature Selection - Classificatore per ogni subset (AC): si riportano i valori dei parametri ottimizzati.



**Figura 4.2:** Migliore coppia Feature Selection - Classificatore per ogni subset (SP): si riportano i valori dei parametri ottimizzati.



Per le 17 migliori coppie (AC e SP), definite nelle figure precedenti, si procede con il testing sul test-set, di cui vengono mostrati i valori di accuratezza nelle Tabelle 4.9 e 4.10.

**Tabella 4.9:** Valori di validation accuracy e test accuracy (AC) per le coppie migliori Feature Selection - Classificatore.

No	FS	Class.	Val_Accuracy	Test_Accuracy
1	Correlation	SVM	86.25%	60.00%
2	ANOVA	SVM	83.57%	80.00%
3	Correlation	SVM	81.25%	75.00%
4	BORUTA	KNN	85.00%	60.00%
5	ANOVA	SVM	74.82%	65.00%
6	BORUTA	GB	83.39%	80.00%
7	BORUTA	KNN	74.82%	55.00%
8	BORUTA	KNN	78.39%	70.00%
9	BORUTA	GB	83.39%	65.00%
10	Correlation	GNB	83.57%	70.00%
11	Correlation	SVM	86.07%	80.00%
12	Correlation	SVM	80.71%	75.00%
13	Correlation	SVM	81.96%	60.00%
14	ANOVA	SVM	83.57%	65.00%
15	BORUTA	SVM	79.64%	80.00%
16	ANOVA	SVM	80.89%	75.00%
17	BORUTA	KNN	86.15%	60.00%

**Tabella 4.10:** Valori di validation accuracy e test accuracy (SP) per le coppie migliori Feature Selection - Classificatore.

No	FS	Class.	Val_Accuracy	Test_Accuracy
1	ANOVA	KNN	81.08%	55.00%
2	Correlation	SVM	80.89%	70.00%
3	BORUTA	KNN	81.07%	85.00%
4	BORUTA	GB	79.82%	80.00%
5	BORUTA	SVM	84.82%	70.00%
6	BORUTA	KNN	77.51%	55.00%
7	BORUTA	GB	84.82%	70.00%
8	ANOVA	SVM	77.32%	50.00%
9	BORUTA	SVM	84.84%	70.00%
10	Correlation	GNB	84.82%	60.00%
11	BORUTA	SVM	88.75%	75.00%
12	Correlation	SVM	84.82%	85.00%
13	BORUTA	KNN	88.39%	85.00%
14	BORUTA	KNN	78.57%	65.00%
15	BORUTA	GB	79.82%	80.00%
16	BORUTA	SVM	81.25%	55.00%
17	ANOVA	SVM	82.14%	60.00%

Nella Tabella 4.11 si riassumono i risultati migliori dell'intero algoritmo di classificazione (sia per AC che per SP), in termini di coppie migliori Feature Selection - Classificatore: si precisano i subset relativi alle coppie migliori, e viene fornita una lista delle relative feature selezionate dallo step di Feature Selection.

**Tabella 4.11:** Migliore coppia FS-Classificatore (AC e SP): si riportano i relativi subset e le feature selezionate.

Dispositivo FS (opt) Modello	AC	SP
	Corr. ( $r = 0.32$ ) SVM	BORUTA (perc = 100) SVM
	'pcm_fftMag_mfcc_sma_2_iqr1_2', 'pcm_fftMag_mfcc_sma_2_iqr1_3', 'pcm_fftMag_mfcc_sma_de_1_percentile1_0', 'pcm_fftMag_spectralFlux_sma_de_kurtosis', 'pcm_fftMag_spectralFlux_sma_kurtosis', 'pcm_fftMag_spectralFlux_sma_linregerrA', 'pcm_fftMag_spectralFlux_sma_gregerrA', 'pcm_fftMag_spectralFlux_sma_skewness', 'pcm_fftMag_spectralKurtosis_sma_de_flatness', 'pcm_fftMag_spectralKurtosis_sma_de_risetime', 'pcm_fftMag_spectralKurtosis_sma_flatness', 'pcm_fftMag_spectralRollOff25_0_sma_lpc0', 'pcm_fftMag_spectralSkewness_sma_de_flatness', 'pcm_fftMag_spectralSkewness_sma_percentile1_0', 'shimmerLocal_sma_flatness'	'pcm_RMSenergy_sma_de_segLenStddev', 'audSpec_Rfilt_sma_15_stddev', 'pcm_fftMag_spectralSkewness_sma_lpc1', 'audSpec_Rfilt_sma_de_7_minPos', 'audSpec_Rfilt_sma_de_7_quartile1', 'audSpec_Rfilt_sma_de_9_lpc0', 'audSpec_Rfilt_sma_de_23_iqr1_2', 'pcm_fftMag_spectralRollOff25_sma_de_downleveltime50', 'voicingFinalUnclipped_sma_pctrange0_1', 'voicingFinalUnclipped_sma_de_quartile1', 'voicingFinalUnclipped_sma_de_percentile1_0', 'voicingFinalUnclipped_sma_de_pctrange0_1', 'audSpec_Rfilt_sma_11_stddevRisingSlope', 'audSpec_Rfilt_sma_11_stddevFallingSlope', 'audSpec_Rfilt_sma_15_peakMeanMeanDist', 'audSpec_Rfilt_sma_15_peakMeanRel', 'pcm_fftMag_spectralFlux_sma_flatness', 'pcm_fftMag_spectralSkewness_sma_stddevFallingSlope', 'voicingFinalUnclipped_sma_stddevRisingSlope'

### 4.1.1 Validazione incrociata

Nella Tabella 4.12 si riassumono le performance della validazione incrociata. Per consentire un confronto con la procedura standard, si riportano i valori di accuratezza sia per il test set costituito da campioni registrati tramite la stessa modalità che nel train set (Test Accuracy), che per il test set costituito da campioni registrati con tecnica diversa (Test accuracy Val. Incr).

**Tabella 4.12:** Valori di train accuracy, test accuracy e test accuracy in seguito a validazione incrociata (Test accuracy Val.Incr): lungo le colonne vengono specificati i subset di partenza, mentre lungo le righe si descrivono le tecniche utilizzate di classificazioni e le metriche di performance calcolate.

No	1		11	
Acquisizione	AC	SP	AC	SP
Classificatore	SVM	SVM	SVM	SVM
Feature selection	Corr (r=0.32)	Corr (r=0.32)	Boruta (perc=100)	Boruta (perc=100)
Train accuracy	85.79%	78.48%	81.99%	95.49%
Test accuracy	85%	70%	65%	80%
Test accuracy Val.Incr	75%	80%	70%	70%

## 4.2 Analisi statistica

Con lo scopo di identificare quali feature estratte dai segnali vocali siano maggiormente invarianti rispetto al sistema di registrazione utilizzato, si è portato avanti il Wilcoxon signed-rank test con tutti i subset (1-17) del dataset LUHS. Affinchè si possa agevolare l'analisi sulle tipologie di features che presentano o meno differenze significative tra AC e SP, si riportano due grafici a barre (Figura 4.3, Figura 4.4):

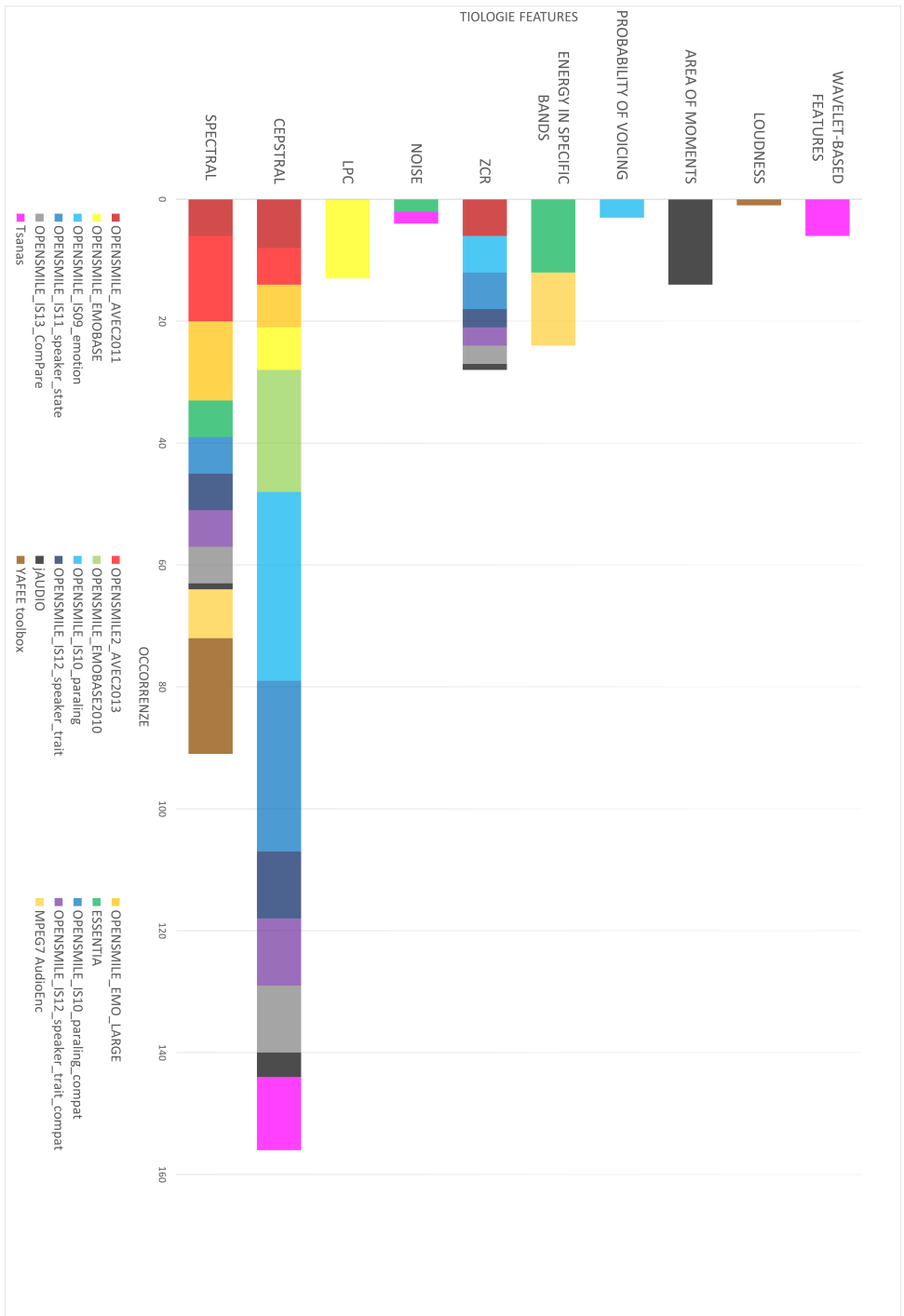
- Uno per le prime 20 features che presentano p-value  $< 0.05$  come risultato del test, e quindi per quelle per cui esiste una differenza significativa tra AC e SP;
- Uno per le ultime 20 features che presentano p-value  $> 0.05$  come risultato del test, e quindi per quelle con cui non si può concludere che ci siano differenze significative tra i due campioni AC e SP.

Prima di passare alla seconda parte dell'analisi statistica, si mostrano nella Figura 4.5 le 20 features del subset di dati *ComPare (13)* [52], che hanno riportato valori di p-value  $< 0.05$  con il dataset LUHS.

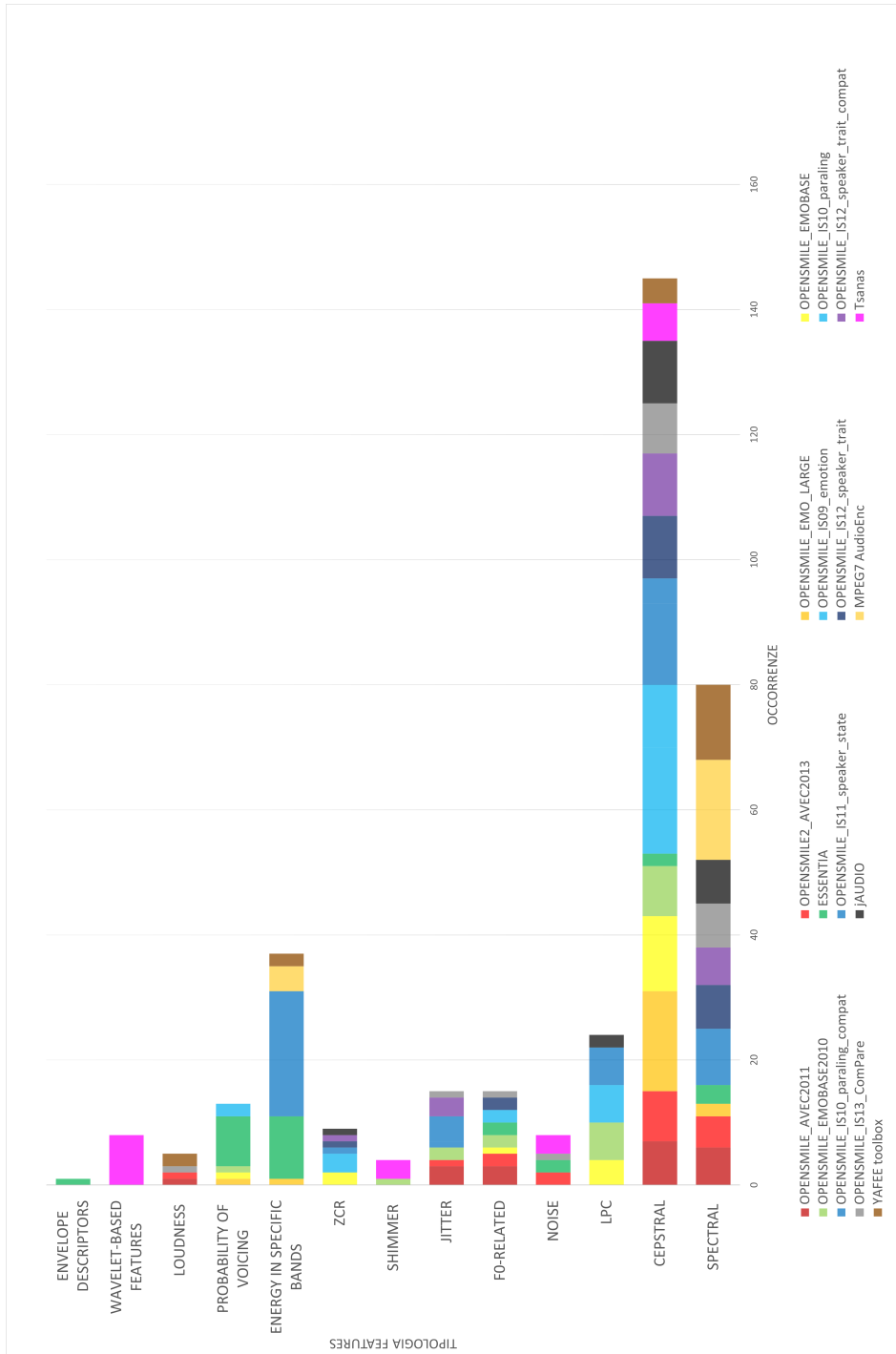
Features (13) p-value < 0.05	P-value	
pcm_zcr_sma_de_iqr2_3	8.58E-15	Zero Crossing Rate
pcm_zcr_sma_de_iqr1_3	6.1E-15	
pcm_zcr_sma_de_posamean	3.07E-15	
pcm_fftMag_spectralVariance_sma_de_quartile1	6.87E-15	Spettrali
pcm_fftMag_spectralVariance_sma_de_iqr1_2	4.17E-15	
pcm_fftMag_spectralVariance_sma_de_percentile99_0	1.75E-15	
pcm_fftMag_spectralVariance_sma_de_iqr2_3	1.25E-15	
pcm_fftMag_spectralVariance_sma_de_iqr1_3	1.08E-15	
pcm_fftMag_spectralVariance_sma_de_quartile3	6.86E-16	MFCC
mfcc_sma_1__percentile1_0	2.52E-15	
mfcc_sma_1__qregc3	1.52E-15	
mfcc_sma_1__linregc2	1.32E-15	
mfcc_sma_5__percentile1_0	1.11E-15	
mfcc_sma_1__percentile99_0	1.87E-16	
mfcc_sma_1__rqmean	1.71E-16	
mfcc_sma_1__amean	1.66E-16	
mfcc_sma_1__quartile2	1.64E-16	
mfcc_sma_1__quartile1	1.62E-16	
mfcc_sma_1__quartile3	1.24E-16	
mfcc_sma_1__peakMeanAbs	6.7E-17	

**Figura 4.5:** Elenco delle prime 20 feature (subset 13) con p-value < 0.05: vengono riportate a destra le famiglie di features a cui appartengono.

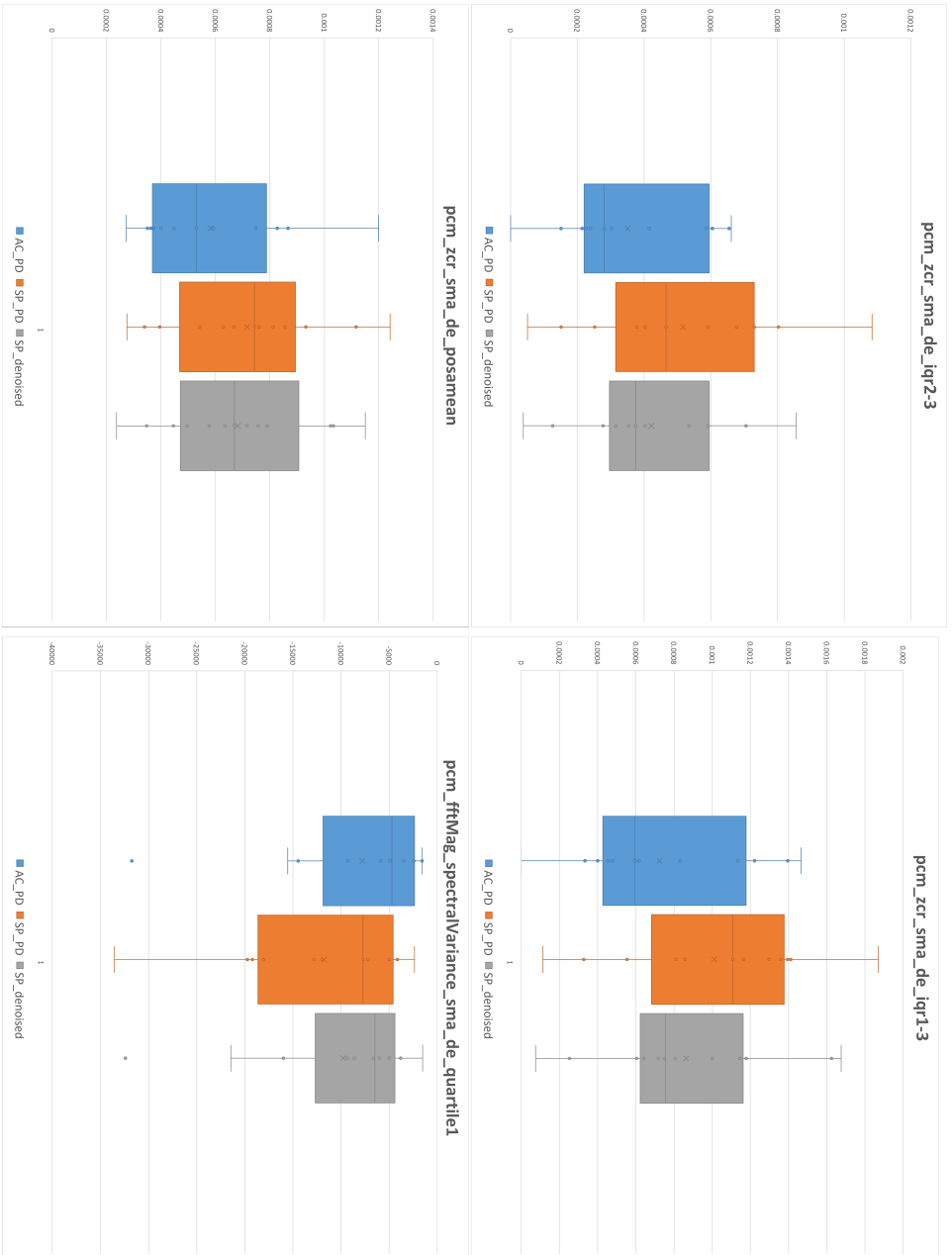
Di queste features, è stata effettuata una analisi statistica grafica attraverso dei Boxplot (Figure 4.6) in cui si mostrano le distribuzioni dei valori sul dataset AAPP. Come ci si aspettava, si riscontrano differenze tra le distribuzioni AC e SP. Andando nel dettaglio, il gruppo di feature in cui si riscontrano maggiori differenze è costituito principalmente da parametri che contengono informazioni circa la banda spettrale associata al rumore: per questo motivo si è deciso di applicare un denoising ai segnali SP, così da valutare la possibilità di mitigare tale fenomeno. In merito al denoising, è stato applicata la *Spectral subtraction* dei segnali SP in *Praat*[39]: si identificano le zone di silenzio tra i vari task e si effettua sottrazione spettrale tra lo spettro del segnale originario e lo spettro delle zone di silenzio. Tale tecnica si basa sul presupposto che nella zona di "silenzio" tutto ciò che viene registrato costituisce "rumore".



**Figura 4.3:** Il grafico riporta il numero di volte in cui una feature con p-value  $< 0.05$  figura tra le prime 20 feature, ordinate in funzione della significatività restituita dal test di Wilcoxon, per ciascuno dei subset di dati in LTHS. La dimensione di ciascun segmento rappresenta il numero di feature relativo alla famiglia di feature e al toolbox considerato.

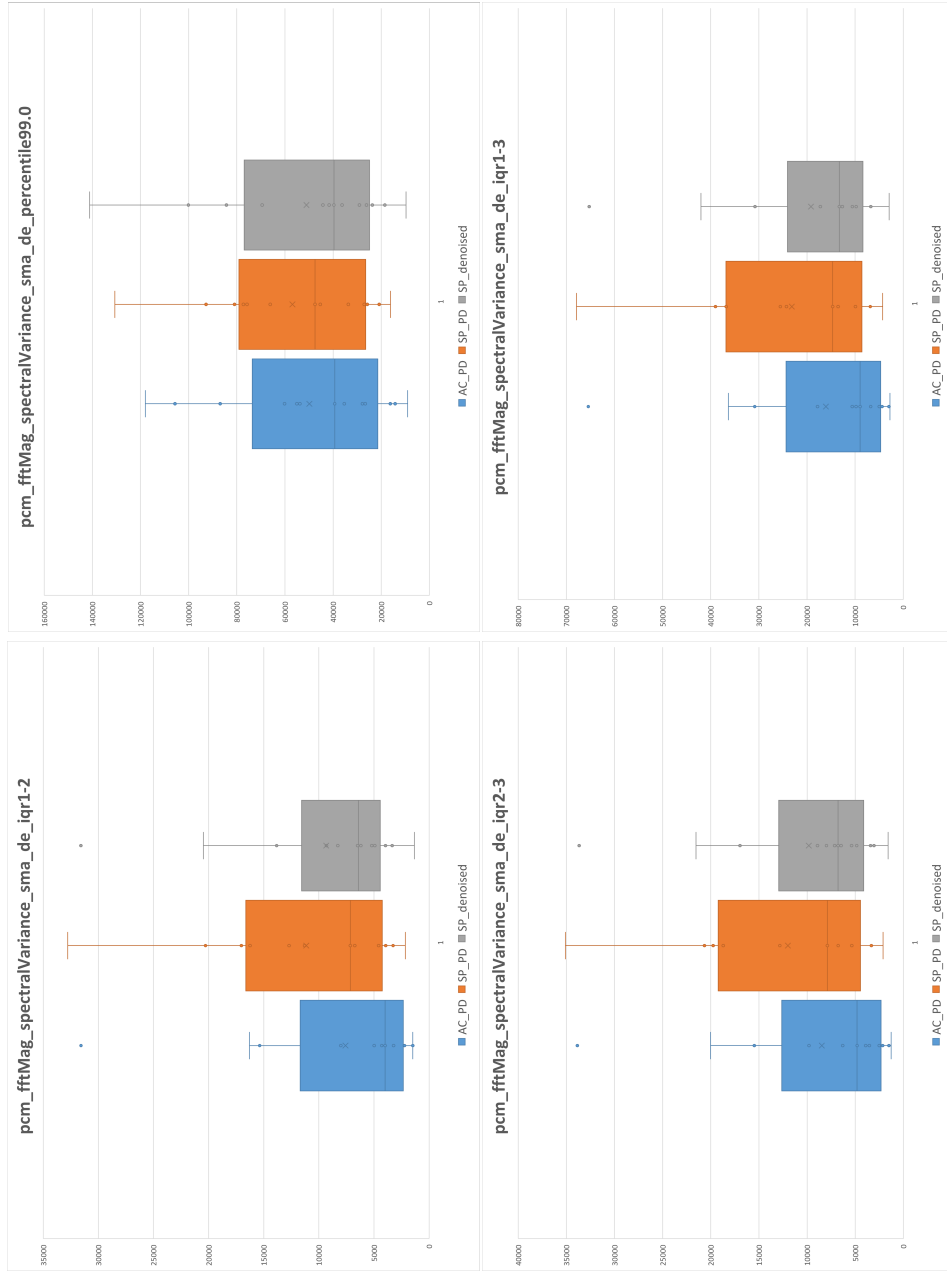


**Figura 4.4:** Il grafico riporta il numero di volte in cui una feature con p-value  $> 0.05$  figura tra le ultime 20 feature, ordinate in funzione della significatività restituita dal test di Wilcoxon, per ciascuno dei subset di dati in LUHS. La dimensione di ciascun segmento rappresenta il numero di feature relativo alla famiglia di feature e al toolbox considerato.

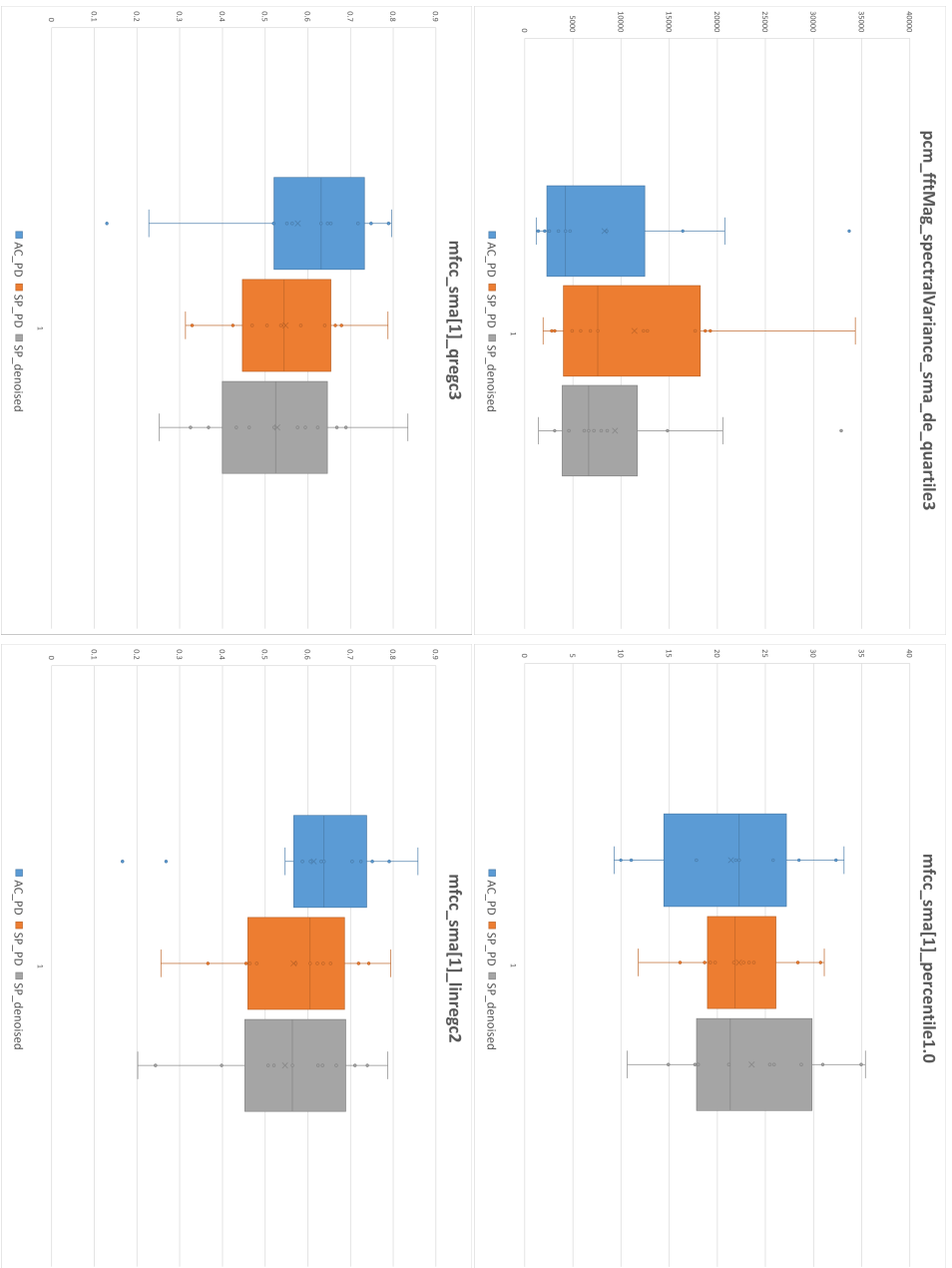


**Figura 4.6:** Boxplot delle distribuzioni di valori delle features del dataset AAPP appartenenti alle registrazioni AC, SP e SP denoised (continua).

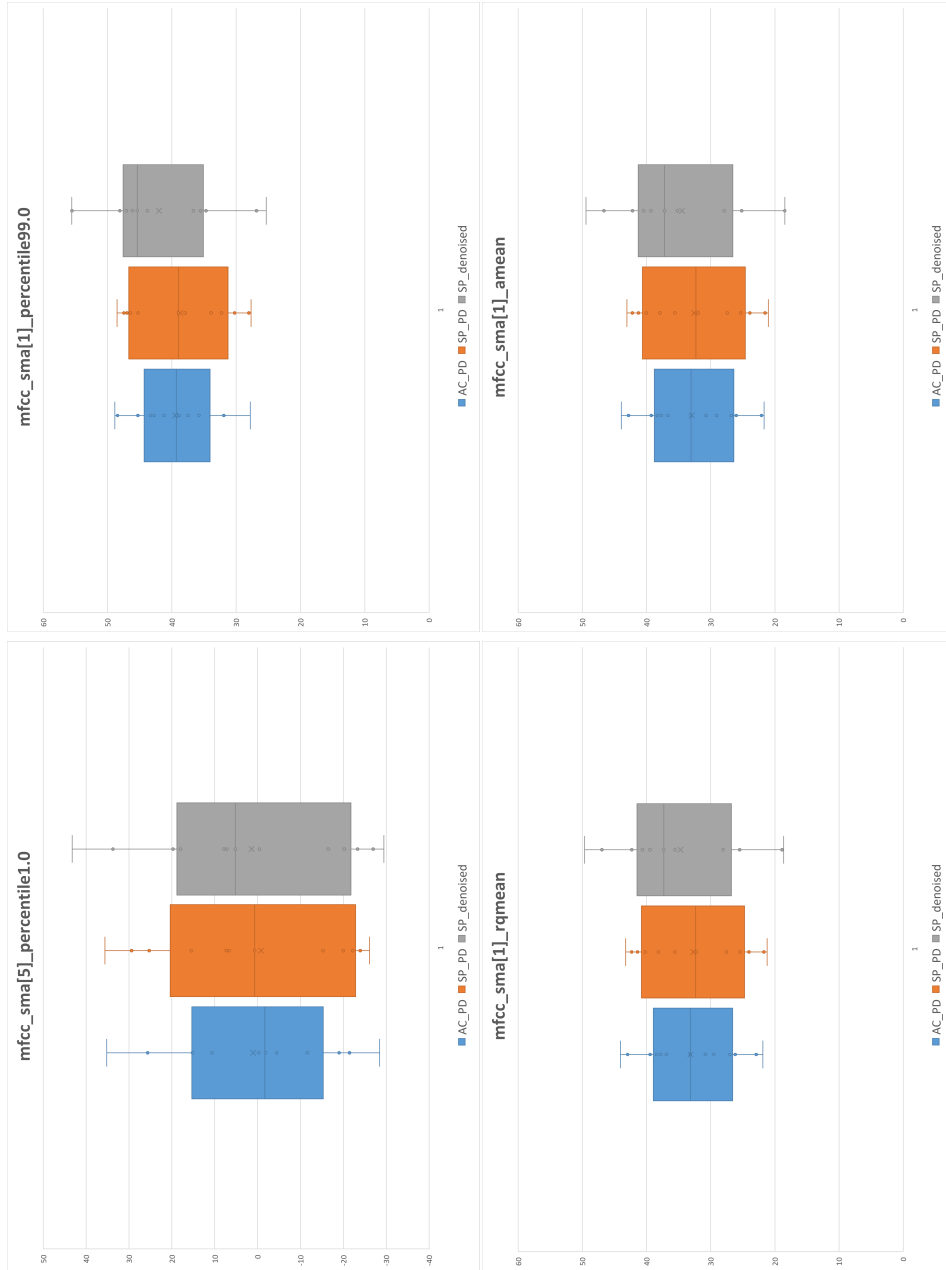




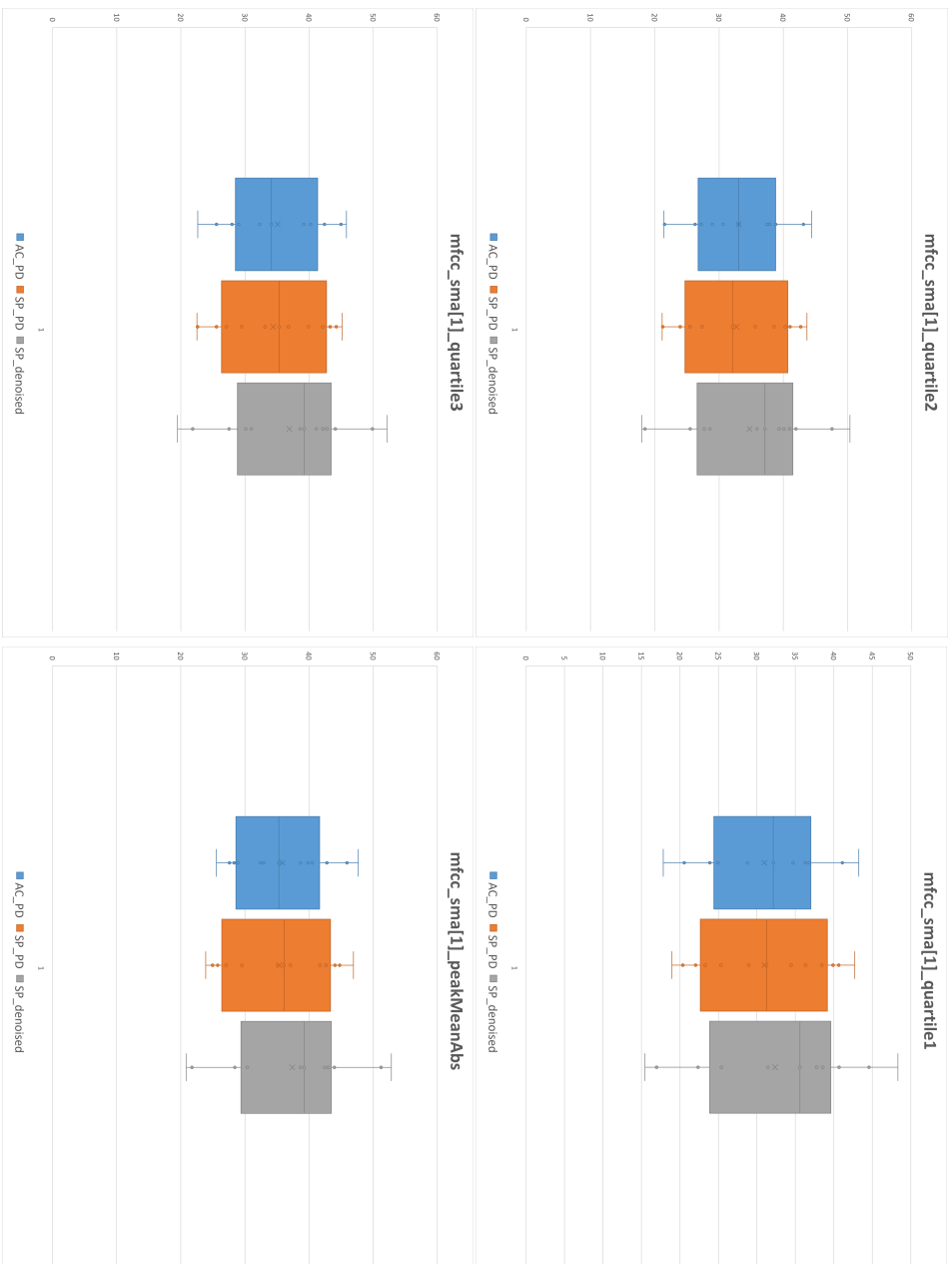
**Figura 4.6:** Boxplot delle distribuzioni di valori delle features del dataset AAPP appartenenti alle registrazioni AC, SP e SP denoised (continua).



**Figura 4.6:** Boxplot delle distribuzioni di valori delle features del dataset APP appartenenti alle registrazioni AC, SP e SP denoised (continua).



**Figura 4.6:** Boxplot delle distribuzioni di valori delle features del dataset AAPP appartenenti alle registrazioni AC, SP e SP denoised (continua).



**Figura 4.6:** Boxplot delle distribuzioni di valori delle features del dataset APP appartenenti alle registrazioni AC, SP e SP denoised.

# Capitolo 5

## Discussioni

### 5.1 Classificazione PD

Dal confronto tra i risultati del validation e del test set (Tabelle 4.9, 4.10) si nota che in alcuni casi è presente overfitting, come ad esempio con il subset 16 (AC) in cui si passa da un valore di validation accuracy pari a 81 a un valore di test accuracy del 55%. Nel dettaglio, i subset che sembrerebbero non presentare overfitting per entrambe le modalità di acquisizione sono i seguenti: 2, 3, 11, 12, 13, 15. Da questi risultati non sembra esserci una correlazione con il modello creato, dato che la procedura è stata uguale per tutti i subset: si può supporre che siano proprio i set di feature e le tecniche di estrazione a generare queste differenze. Benchè si apprezzi una parziale riduzione di performance passando da validation a test set, la maggior parte dei subset non presenta overfitting.

I risultati ottenuti dall’algoritmo di classificazione sono stati confrontati con quelli ottenuti dallo studio di *Ali et al.* [14]: la scelta di tale studio è stata dettata tanto dall’utilizzo dello stesso dataset (LUHS), quanto dal fatto che sia l’unico studio trovato in letteratura che riporta in maniera dettagliata le prestazioni per ciascun modello costruito.

Nelle Tabelle 5.1 e 5.2 si riportano a confronto i valori di test accuracy massimi raggiunti nel lavoro di tesi e nello studio, con il task di vocale sostenuta /a/.

Ai fini del confronto vale la pena notare che in [14] non viene applicata nessuna tecnica di feature selection e l’intero set di feature viene utilizzato come input del classificatore. Inoltre, non vengono riportati i risultati sul train e validation set, il che non consente di valutare un eventuale overfitting.

**Tabella 5.1:** AC: Valori a confronto di test accuracy tra l’algoritmo di classificazione sviluppato nel lavoro di tesi e quello dello studio [14]. Si evidenziano i set di dati con cui si raggiungono valori più alti nel lavoro di tesi rispetto allo studio. FS\_T: Feature Selection utilizzata nel lavoro di tesi; FS\_A: Feature Selection utilizzata nello studio; Class\_T: Classificatore utilizzato nel lavoro di tesi; Class\_A: Classificatore utilizzato nello studio; Test\_Accuracy\_T: Accuratezza raggiunta nel lavoro di tesi; Test\_Accuracy\_A: Accuratezza raggiunta nello studio.

No	FS_T	FS_A	Class_T	Class_A	Test_Accuracy_T	Test_Accuracy_A
1	Correlation	-	SVM	LDA	60.00%	66.66%
2	<b>ANOVA</b>	-	<b>SVM</b>	<b>LDA</b>	<b>80.00%</b>	<b>66.66%</b>
3	<b>Correlation</b>	-	<b>SVM</b>	<b>LDA</b>	<b>75.00%</b>	<b>74.66%</b>
4	BORUTA	-	KNN	LDA	60.00%	73.33%
5	ANOVA	-	SVM	LDA	65.00%	72.00%
6	<b>BORUTA</b>	-	<b>GB</b>	<b>LDA</b>	<b>80.00%</b>	<b>76.00%</b>
7	BORUTA	-	KNN	SVM	55.00%	65.33%
8	BORUTA	-	KNN	LDA	70.00%	70.66%
9	BORUTA	-	GB	KNN	65.00%	66.66%
10	<b>Correlation</b>	-	<b>GNB</b>	<b>KNN</b>	<b>70.00%</b>	<b>66.66%</b>
11	<b>Correlation</b>	-	<b>SVM</b>	<b>LDA</b>	<b>80.00%</b>	<b>69.33%</b>
12	<b>Correlation</b>	-	<b>SVM</b>	<b>LDA</b>	<b>75.00%</b>	<b>68.00%</b>
13	Correlation	-	SVM	SVM	60.00%	65.33%
14	ANOVA	-	SVM	LDA	65.00%	74.66%
15	<b>BORUTA</b>	-	<b>SVM</b>	<b>GNB</b>	<b>80.00%</b>	<b>72.00%</b>
16	<b>ANOVA</b>	-	<b>SVM</b>	<b>GNB</b>	<b>75.00%</b>	<b>68.00%</b>
17	BORUTA	-	KNN	LDA	60.00%	84.00%

Dai confronti si evidenzia che per le registrazioni SP, attraverso l’algoritmo di classificazione delineato nel lavoro di tesi, si raggiungono valori di accuratezza sul test set maggiori rispetto allo studio per più della metà dei set di dati, precisamente per 12: nel caso dei subset 3,12 e 13 viene riportata una accuratezza massima pari all’80%, contro il 74.66% raggiunto dallo studio. Per le registrazioni AC invece, valori maggiori di accuratezza si registrano su 8 set di dati, raggiungendo l’80% di test accuracy.

Da tali risultati si deduce che l’applicazione della *Feature Selection*, nel caso di SP, ha permesso di raggiungere risultati migliori, consentendo un allenamento con feature ad hoc per i diversi subset.

Altro fattore che con buona probabilità ha contribuito a tali risultati è stata la scelta della divisione training e test set 80-20%, contro una divisione 70-30% adottata da Ali et al. Questo è probabilmente legato al numero limitato di registrazioni del dataset LUHS: scegliendo un numero maggiore di registrazioni con cui allenare i classificatori, è stato possibile raggiungere risultati migliori. Infatti, a differenza di quanto effettuato in [14], si è ritenuto opportuno non ridurre la

**Tabella 5.2:** SP: Valori a confronto di test accuracy tra l’algoritmo di classificazione sviluppato nel lavoro di tesi e quello dello studio [14]. Si evidenziano i set di dati con cui si raggiungono valori più alti nel lavoro di tesi rispetto allo studio. FS\_T: Feature Selection utilizzata nel lavoro di tesi; FS\_A: Feature Selection utilizzata nello studio; Class\_T: Classificatore utilizzato nel lavoro di tesi; Class\_A: Classificatore utilizzato nello studio; Test\_Accuracy\_T: Accuratezza raggiunta nel lavoro di tesi; Test\_Accuracy\_A: Accuratezza raggiunta nello studio.

No	FS_T	FS_A	Class_T	Class_A	Test_Accuracy_T	Test_Accuracy_A
1	ANOVA	-	KNN	KNN	55.00%	68.00%
2	<b>Correlation</b>	-	<b>SVM</b>	<b>LDA</b>	<b>70.00%</b>	<b>73.33%</b>
3	<b>BORUTA</b>	-	<b>KNN</b>	<b>LDA</b>	<b>85.00%</b>	<b>74.66%</b>
4	<b>BORUTA</b>	-	<b>GB</b>	<b>LDA</b>	<b>80.00%</b>	<b>72.00%</b>
5	<b>BORUTA</b>	-	<b>SVM</b>	<b>SVM</b>	<b>70.00%</b>	<b>65.33%</b>
6	<b>BORUTA</b>	-	<b>KNN</b>	<b>LDA</b>	<b>55.00%</b>	<b>74.66%</b>
7	<b>BORUTA</b>	-	<b>GB</b>	<b>KNN</b>	<b>70.00%</b>	<b>69.33%</b>
8	ANOVA	-	SVM	LDA	50.00%	69.33%
9	BORUTA	-	SVM	LDA	70.00%	73.33%
10	<b>Correlation</b>	-	<b>GNB</b>	<b>KNN</b>	<b>60.00%</b>	<b>66.66%</b>
11	<b>BORUTA</b>	-	<b>SVM</b>	<b>SVM</b>	<b>75.00%</b>	<b>65.33%</b>
12	<b>Correlation</b>	-	<b>SVM</b>	<b>SVM</b>	<b>85.00%</b>	<b>65.33%</b>
13	<b>BORUTA</b>	-	<b>KNN</b>	<b>LDA</b>	<b>85.00%</b>	<b>68.00%</b>
14	BORUTA	-	KNN	LDA	65.00%	73.33%
15	<b>BORUTA</b>	-	<b>GB</b>	<b>GNB</b>	<b>80.00%</b>	<b>65.33%</b>
16	<b>BORUTA</b>	-	<b>SVM</b>	<b>GNB</b>	<b>55.00%</b>	<b>69.33%</b>
17	ANOVA	-	SVM	GNB	60.00%	66.66%

numerosità del dataset al di sotto dell’80% per evitare di allenare un modello eccessivamente adattato agli specifici esempi forniti, producendo un elevato effetto di overfitting. Al contempo, la presenza di un test set di dimensioni ridotte potrebbe aver determinato una parziale riduzione della stima delle performance nel modello (a causa della limitata numerosità del dataset).

Dal confronto tra i risultati ottenuti dalle registrazioni AC e SP, riassunti in Tabella 5.3, si nota come non siano presenti evidenti differenze tra i valori di accuratezza, prova preliminare del fatto che l’algoritmo di classificazione sviluppato raggiunge risultati ottimi sia con segnali derivanti da microfono che da smartphone. C’è da evidenziare la presenza di migliori performance nei subset SP rispetto ad AC. Mettendo a sistema tale evidenza con i risultati del confronto con lo studio [14], da cui era emersa una generale superiorità dei modelli proposti in questa tesi solo per il dataset SP, si deduce la necessità di preferire, per la modalità AC, l’assenza di uno step di feature selection.

**Tabella 5.3:** Valori a confronto di test accuracy ottenuti con registrazioni AC e SP. Si evidenziano i valori di accuratezza più alti.

No	FS_AC	FS_SP	Class_AC	Class_SP	Test_Acc_AC	Test_Acc_SP
1	Correlation	ANOVA	SVM	KNN	60.00%	55.00%
2	ANOVA	Correlation	SVM	SVM	<b>80.00%</b>	70.00%
3	Correlation	BORUTA	SVM	KNN	75.00%	<b>85.00%</b>
4	BORUTA	BORUTA	KNN	GB	60.00%	80.00%
5	ANOVA	BORUTA	SVM	SVM	65.00%	70.00%
6	BORUTA	BORUTA	GB	KNN	<b>80.00%</b>	55.00%
7	BORUTA	BORUTA	KNN	GB	55.00%	70.00%
8	BORUTA	ANOVA	KNN	SVM	70.00%	50.00%
9	BORUTA	BORUTA	GB	SVM	65.00%	70.00%
10	Correlation	Correlation	GNB	GNB	70.00%	60.00%
11	Correlation	BORUTA	SVM	SVM	<b>80.00%</b>	75.00%
12	Correlation	Correlation	SVM	SVM	75.00%	<b>85.00%</b>
13	Correlation	BORUTA	SVM	KNN	60.00%	<b>85.00%</b>
14	ANOVA	BORUTA	SVM	KNN	65.00%	65.00%
15	BORUTA	BORUTA	SVM	GB	<b>80.00%</b>	80.00%
16	ANOVA	BORUTA	SVM	SVM	75.00%	55.00%
17	BORUTA	ANOVA	KNN	SVM	60.00%	60.00%

### 5.1.1 Validazione incrociata

Per approfondire l’analisi sulla indipendenza dell’algoritmo di classificazione dalla modalità di acquisizione, si passa ai risultati ottenuti dallo step di *validazione incrociata*, riassunti in Tabella 4.12. Sono presenti delle parziali differenze tra i valori di accuratezza di AC e SP, che però sono state mitigate dall’ottimizzazione della Feature Selection, che consente dunque di migliorare i risultati. Ciò si nota sia dai valori sul Train che sul Test set; infatti, per il subset 1, ottimizzato per le registrazioni AC, si riscontrano valori di accuratezza maggiori per AC rispetto ad SP, prova di un corretto funzionamento dell’ottimizzazione. Lo stesso si verifica per



il subset 11, questa volta ottimizzato per le registrazioni SP. Passando al confronto incrociato:

- Prendendo in considerazione il subset ottimizzato per AC (1), preliminarmente non si nota un evidente overfitting, dal momento che le performance da train a test si abbassano di meno dell'1%.
- Considerando invece le registrazioni SP sul subset 1, le performance risultano abbassarsi, dal momento che si è utilizzato un modello non ottimizzato per questo set di dati. Analizzando invece il risultato di Test accuracy Val.Incr, si riscontra un aumento dal 70% all'80%: questo perchè il set di feature utilizzato nella validazione incrociata è effettivamente quello ottimizzato per tale subset. Complessivamente, si riscontra una riduzione delle performance del 10% testando sugli smartphone un modello costruito su dati registrati tramite microfoni professionali.
- Prendendo in considerazione il subset ottimizzato per SP (11), anche in questo caso non si nota un evidente overfitting passando dai valori di train a quelli di test.
- Considerando invece le registrazioni AC sul subset 11, le performance risultano ancora una volta abbassarsi, dal momento che si è utilizzato un modello non ottimizzato per questo set di dati. Analizzando invece il risultato di Test accuracy Val.Incr, si riscontra un aumento dal 65% al 70%: questo perchè il modello utilizzato nella validazione incrociata è effettivamente quello ottimizzato per tale subset. Anche in questo caso si verifica una riduzione del 10% testando sui microfoni professionali un modello costruito su dati registrati tramite smartphone.

## 5.2 Analisi statistica

Ciò che emerge dalle Figure 4.3, 4.4, può essere schematizzato nei seguenti punti:

- *F0-related* e *Jitter* sono presenti esclusivamente tra la popolazione con  $p\text{-value} > 0.05$ , e in particolare si ritrovano rispettivamente in 8 e 6 subset differenti; da ciò si deduce che sia con F0 che con Jitter non si riscontrano differenze evidenti tra i valori derivanti dalle registrazioni AC e SP. Ciò trova conferma in vari studi in cui l'obiettivo è lo stesso, ovvero capire quali features presentano valori paragonabili tra registrazioni AC e SP [38, 40, 41].
- Le varianti di *Shimmer* sono presenti esclusivamente nella popolazione con  $p\text{-value} > 0.05$ , ma con occorrenza bassa, poichè si ritrovano solo in 2 toolboxes; anche se meno evidente rispetto a F0 e Jitter, anche le varianti di Shimmer non risultano differire tra AC e SP;

- Anche gli *Envelope Descriptors* sono presenti esclusivamente tra le features con  $p\text{-value} > 0.05$ . C'è da precisare che queste tipologie di features, insieme a *Wavelet based features* e *Area of moments* sono state estratte ciascuna da un solo toolbox, per cui i risultati necessitano ulteriore validazione;
- Le features *Area of moments* compaiono esclusivamente tra le feature con  $p\text{-value} < 0.05$ , per cui si può supporre che presentino grosse differenze tra AC e SP;
- *Cepstral* e *Spectral* necessitano di una analisi più approfondita, in quanto presenti in egual modo in entrambe le popolazioni.

Nelle Figure 5.1, 5.2 si mostrano in modo approfondito le occorrenze dei vari descrittori statistici della famiglia *Spectral*, sia tra le features con  $p\text{-value} < 0.05$  che tra quelle con  $p\text{-value} > 0.05$ .

Dai due grafici è evidente una sostanziale differenza nelle occorrenze dei descrittori statistici *Variance* e *Flatness*: la prima compare solo tra le feature con  $p\text{-value} < 0.05$ , segno di una maggiore dipendenza dalla modalità di acquisizione, mentre il secondo solo tra le features con  $p\text{-value} > 0.05$ .

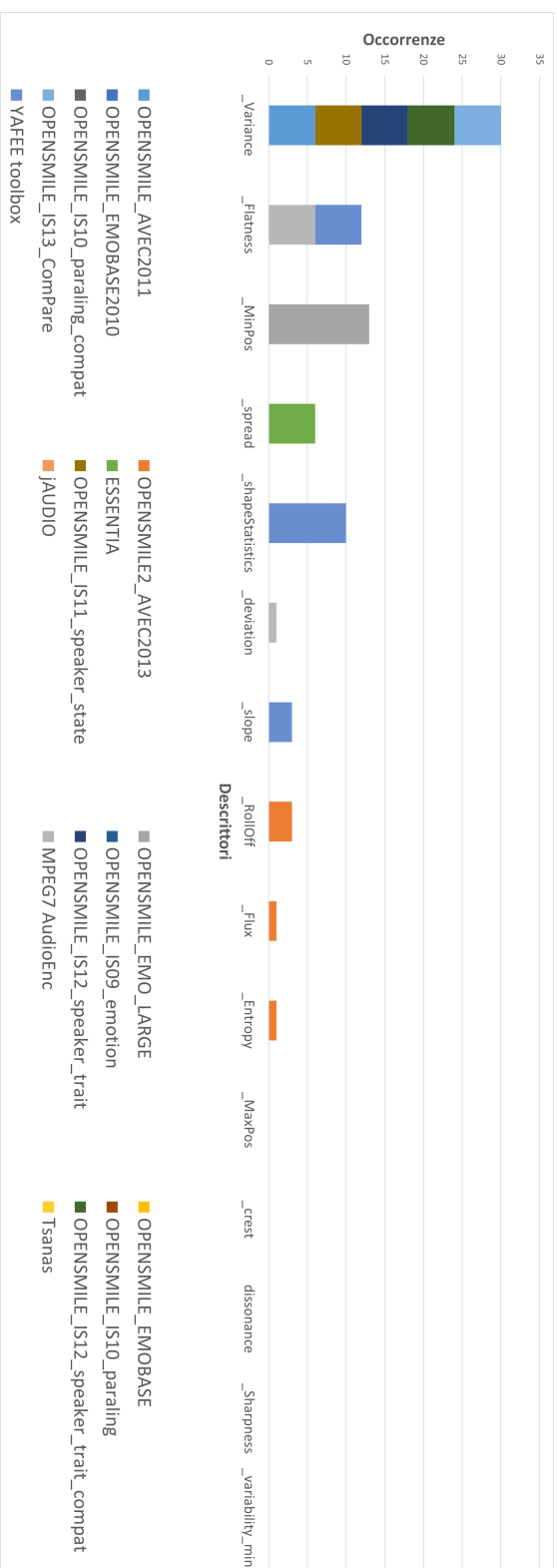
Come per la famiglia *Spectral*, anche per le *Cepstral* si mostrano nelle Figure 5.3, 5.4 le occorrenze dei vari gradi di MFCC. C'è da precisare che questa analisi presenta il limite che in alcuni toolbox il numero di coefficienti MFCC utilizzato è 25 e non 13, per cui la corrispondenza non è univoca; ad ogni modo, i toolbox che utilizzano 25 coefficienti sono la minoranza.

Ciò che ne risulta dal confronto tra i due grafici è che il primo coefficiente (MFCC1) mostra maggiori differenze tra i dispositivi, in quanto la sua occorrenza tra le features con  $p\text{-value} < 0.05$  è di 81 unità, contro i 12 con  $p\text{-value} > 0.05$ . Tale fenomeno può essere dovuto alla maggiore presenza di rumore di fondo delle registrazioni tramite smartphone, considerando che i primi coefficienti di MFCC parametrizzano maggiormente la banda del rumore: da qui la scelta di effettuare il denoising sulle registrazioni SP del dataset AAPP, di modo tale da verificare un eventuale mitigazione del rumore.

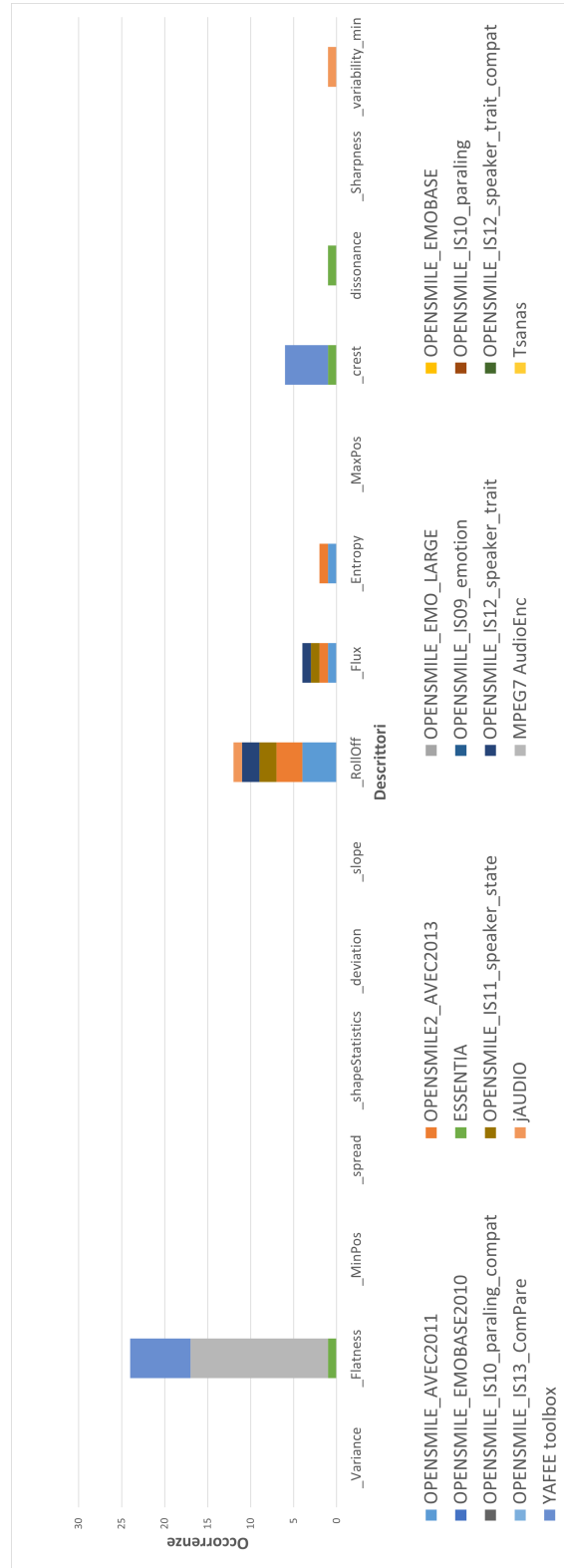
A questo proposito, ciò che emerge dall'analisi dei Boxplot (Figura 4.6) si può riassumere nei seguenti punti:

- Le distribuzioni di valori SP denoised, per la famiglia di feature *ZCR*, si ridistribuiscono sovrapponendosi maggiormente alle distribuzioni AC, per cui si deduce che il denoising ha effettivamente mitigato gli effetti dovuti al rumore;
- Allo stesso modo, anche per le feature appartenenti al gruppo *Spectral* il denoising ha funzionato nell'intento di mitigare il rumore del segnale, dal momento che le distribuzioni si avvicinano a quelle di AC.
- Non si può dedurre lo stesso invece dalle distribuzioni degli MFCC (*Cepstral*). Per queste feature il denoising non apporta alcuna ridistribuzione dei valori.

Complessivamente il denoising sembra funzionare fintanto che si valutano semplici coefficienti spettrali, ma non è più sufficiente quando si valutano parametri più complessi, come gli MFCC.



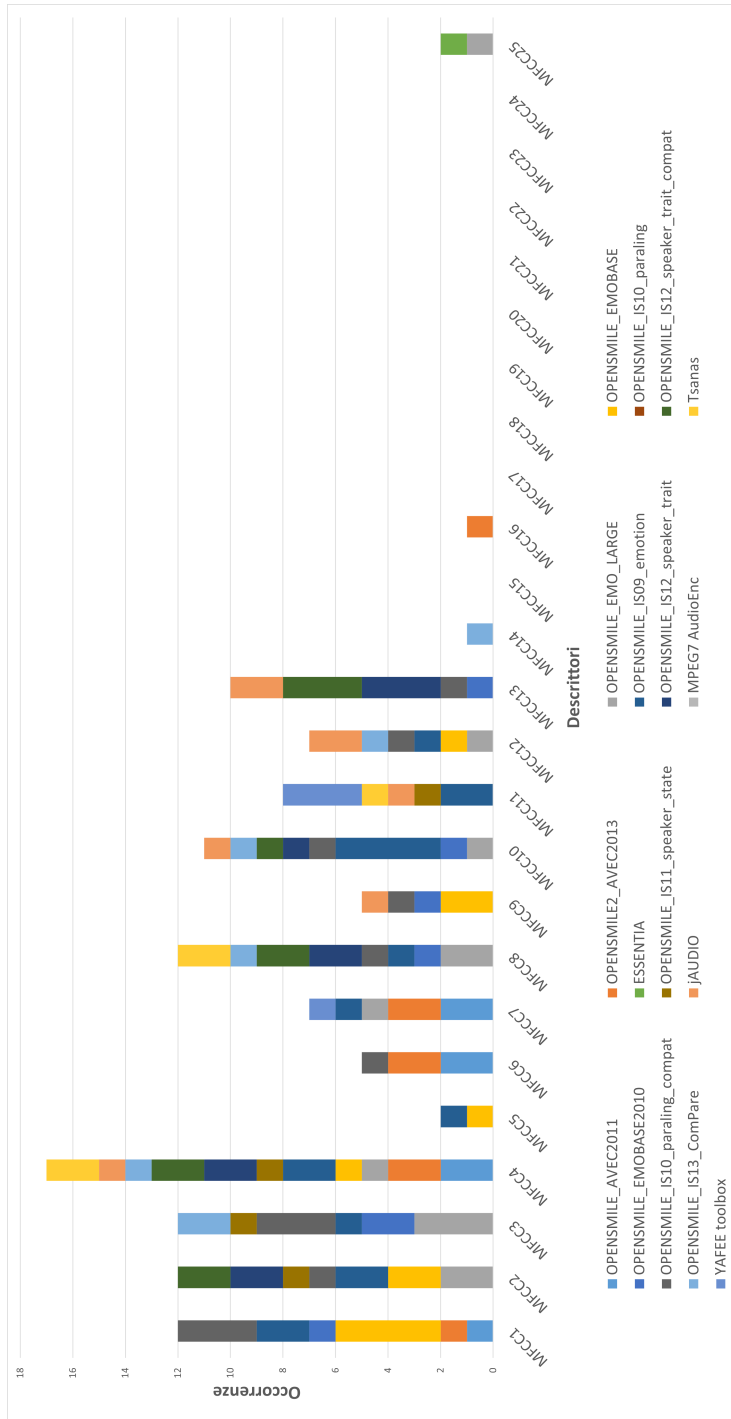
**Figura 5.1:** Il grafico riporta il numero di volte in cui i descriptori delle feature *Spectral* con p-value < 0.05 figurano tra le prime 20 feature, ordinate in funzione della significatività restituita dal test di Wilcoxon, per ciascuno dei subset di dati in LHS.



**Figura 5.2:** Il grafico riporta il numero di volte in cui i descrittori delle feature *Spectral* con  $p\text{-value} > 0.05$  figurano tra le ultime 20 feature, ordinate in funzione della significatività restituita dal test di Wilcoxon, per ciascuno dei subset di dati in LUHS.



**Figura 5.3:** Il grafico riporta il numero di volte in cui i descrittori delle feature *MFCC* con *p-value*  $< 0.05$  figurano tra le prime 20 feature, ordinate in funzione della significatività restituita dal test di Wilcoxon, per ciascuno dei subset di dati in LHS.



**Figura 5.4:** Il grafico riporta il numero di volte in cui i descrittori delle feature *MFCC* con  $p\text{-value} > 0.05$  figurano tra le ultime 20 feature, ordinate in funzione della significatività restituita dal test di Wilcoxon, per ciascuno dei subset di dati in LUHS.





## Capitolo 6

# Conclusioni

L'obiettivo posto da questo lavoro di tesi è la validazione delle registrazioni vocali acquisite con smartphone di pazienti parkinsoniani, e dunque il confronto delle caratteristiche acustiche rappresentative ottenute parallelamente attraverso uno smartphone e un microfono professionale. Tale obiettivo è stato raggiunto seguendo due strade:

- Attraverso la creazione di un algoritmo di rilevamento della malattia di Parkinson ad hoc per registrazioni SP e AC, con uno step conclusivo di *validazione incrociata*: ciò che si può concludere è che è possibile utilizzare entrambe le tipologie di registrazioni ma è necessario costruire modelli ad hoc che abbiano feature e algoritmi adatti. In alternativa è possibile mitigare parzialmente l'influenza del dispositivo utilizzando tecniche di denoising ma, come visto, questo copre solo una parte delle alterazioni, per cui è necessario investigare ulteriormente l'argomento in futuro;
- Attraverso una analisi statistica delle feature estratte dai vari toolbox è emerso che *F0* e *Jitter* sono le feature che più di tutte non mostrano differenze significative tra SP e AC. Ciò risulta meno evidente per le *varianti di Shimmer*, che comunque possono essere considerate come feature invariante tra SP e AC. Bisogna invece prestare attenzione con le famiglie Cepstral e Spectral, e in generale con tutte quelle famiglie di features che risultano dipendere maggiormente dal rumore di fondo. A tal proposito è stato applicato un denoising di tipo *Spectral Subtraction* alle registrazioni SP: questa ha funzionato nell'intento di mitigare il rumore di fondo con le famiglie *Spectral* e *ZCR*, feature che presentavano una preliminare differenza tra le due modalità di acquisizione. Ciò non è avvenuto invece con gli MFCC relativi ai primi coefficienti.

Un eventuale miglioramento futuro potrebbe essere l'inclusione di ulteriori features nell'analisi statistica, di modo tale da comprendere per ogni famiglia dettagliatamente quali descrittori sono maggiormente invariati rispetto alla modalità di acquisizione; per attuare ciò bisognerebbe ampliare il dataset di partenza, rendendo l'analisi statistica sempre più affidabile. Per quanto riguarda invece l'algoritmo di classificazione, le prestazioni potrebbero migliorare ulteriormente considerando altri task di registrazione oltre che la fonazione sostenuta: questo consentirebbe l'inclusione di altre features che forniscono informazioni circa gli aspetti articolatori e prosodici del PD.

Dal confronto in termini di performance tra AC e SP è stata dedotta la necessità di preferire, per la modalità AC, l'assenza di uno step di feature selection. In alternativa, al fine di evitare di sovradimensionare il problema, un miglioramento futuro potrebbe essere l'utilizzo di algoritmi alternativi di selezione delle feature. Al contrario, i modelli realizzati sembrano fornire risultati estremamente promettenti per l'applicazione con registrazioni di qualità sub-ottima.

# Bibliografia

- [1] Ian McLoughlin. *Applied Speech and Audio Processing: With Matlab Examples*. Cambridge University Press, 2009. DOI: 10.1017/CB09780511609640 (cit. a p. 1).
- [2] Meyer TK. «The larynx for neurologists». In: (nov. 2009). DOI: 10.1097/NRL.0b013e3181b1cde5 (cit. a p. 1).
- [3] Jankovic J. «Parkinson's disease: clinical features and diagnosis». In: (2008). DOI: 10.1136/jnnp.2007.131045 (cit. alle pp. 2, 3).
- [4] Lorraine V Kalia e Anthony E Lang. «Parkinson's disease». In: *The Lancet* 386.9996 (2015), pp. 896–912. ISSN: 0140-6736. DOI: [https://doi.org/10.1016/S0140-6736\(14\)61393-3](https://doi.org/10.1016/S0140-6736(14)61393-3). URL: <https://www.sciencedirect.com/science/article/pii/S0140673614613933> (cit. a p. 2).
- [5] Laureano Moro-Velazquez, Jorge A. Gomez-Garcia, Julian D. Arias-Londoño, Najim Dehak e Juan I. Godino-Llorente. «Advances in Parkinson's Disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects». In: *Biomedical Signal Processing and Control* 66 (2021), p. 102418. ISSN: 1746-8094. DOI: <https://doi.org/10.1016/j.bspc.2021.102418>. URL: <https://www.sciencedirect.com/science/article/pii/S174680942100015X> (cit. alle pp. 3, 5).
- [6] Giovanni Defazio, Marta Guerrieri, Daniele Liuzzi, Angelo Fabio Gigante e Vincenzo di Nicola. «Assessment of voice and speech symptoms in early Parkinson's disease by the Robertson dysarthria profile». In: *Neurological Sciences* 37 (3 2016). ISSN: 15903478. DOI: 10.1007/s10072-015-2422-8 (cit. a p. 3).
- [7] Mackenzie C. «Dysarthria in stroke: a narrative review of its description and the outcome of intervention». In: (apr. 2011). DOI: 10.3109/17549507.2011.52494 (cit. a p. 3).

- 
- [8] Jan Rusz, Jan Hlavnicka, Tereza Tykalova, Michal Novotny, Petr Dusek, Karel Sonka e Evzen Ruzicka. «Smartphone Allows Capture of Speech Abnormalities Associated with High Risk of Developing Parkinson’s Disease». In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26 (8 2018). ISSN: 15344320. DOI: 10.1109/TNSRE.2018.2851787 (cit. alle pp. 3, 7, 13–15).
- [9] L. Brabenec, J. Mekyska, Z. Galaz e Irena Rektorova. «Speech disorders in Parkinson’s disease: early diagnostics and effects of medication and brain stimulation». In: *Journal of Neural Transmission* 124 (3 2017). ISSN: 14351463. DOI: 10.1007/s00702-017-1676-0 (cit. a p. 3).
- [10] Jan Hlavnicka, Roman Cmejla, Tereza Tykalová, Karel Šonka, Evzen Ruzicka e Jan Rusz. «Automated analysis of connected speech reveals early biomarkers of Parkinson’s disease in patients with rapid eye movement sleep behaviour disorder». In: *Scientific Reports* 7 (1 2017). ISSN: 20452322. DOI: 10.1038/s41598-017-00047-5 (cit. a p. 3).
- [11] Adam P. Vogel, Kristin M. Rosen, Angela T. Morgan e Sheena Reilly. «Comparability of modern recording devices for speech analysis: Smartphone, landline, laptop, and hard disc recorder». In: *Folia Phoniatrica et Logopaedica* 66 (6 2014). ISSN: 14219972. DOI: 10.1159/000368227 (cit. a p. 3).
- [12] Max A. Little, Patrick E. McSharry, Eric J. Hunter, Jennifer Spielman e Lorraine O. Ramig. «Suitability of dysphonia measurements for telemonitoring of Parkinson’s disease». In: *IEEE Transactions on Biomedical Engineering* 56 (4 2009). ISSN: 00189294. DOI: 10.1109/TBME.2008.2005954 (cit. a p. 6).
- [13] Alex S. Ozbolt, Laureano Moro-Velazquez, Ioan Lina, Ankur A. Butala e Najim Dehak. «Things to Consider When Automatically Detecting Parkinson’s Disease Using the Phonation of Sustained Vowels: Analysis of Methodological Issues». In: *Applied Sciences* 12.3 (2022). DOI: 10.3390/app12030991 (cit. a p. 6).
- [14] «MMDD-Ensemble: A Multimodal Data–Driven Ensemble Approach for Parkinson’s Disease Detection». In: *Frontiers in Neuroscience* 15 (2021). ISSN: 1662453X. DOI: 10.3389/fnins.2021.754058 (cit. alle pp. 6, 7, 53–56).
- [15] Indira Rustempasic e Mehmet Can. «Diagnosis of Parkinson’s Disease using Fuzzy C-Means Clustering and Pattern Recognition». In: *Southeast Europe Journal of Soft Computing* 2 (1 2013). DOI: 10.21533/scjournal.v2i1.44 (cit. a p. 6).
- [16] Mohammad Shahbakhi, Danial Taheri Far e Ehsan Tahami. «Speech Analysis for Diagnosis of Parkinson’s Disease Using Genetic Algorithm and Support Vector Machine». In: *Journal of Biomedical Science and Engineering* 07 (04 2014). ISSN: 1937-6871. DOI: 10.4236/jbise.2014.74019 (cit. a p. 6).

- 
- [17] Haydar Ozkan. «A comparison of classification methods for telediagnosis of Parkinson's disease». In: *Entropy* 18 (4 2016). ISSN: 10994300. DOI: 10.3390/e18040115 (cit. a p. 6).
- [18] Liaqat Ali, Ce Zhu, Zhonghao Zhang e Yipeng Liu. «Automated Detection of Parkinson's Disease Based on Multiple Types of Sustained Phonations Using Linear Discriminant Analysis and Genetically Optimized Neural Network». In: *IEEE Journal of Translational Engineering in Health and Medicine* 7 (2019). ISSN: 21682372. DOI: 10.1109/JTEHM.2019.2940900 (cit. a p. 6).
- [19] Emre Avuçlu e Abdullah Elen. «Evaluation of train and test performance of machine learning algorithms and Parkinson diagnosis with statistical measurements». In: *Medical and Biological Engineering and Computing* 58 (11 2020). ISSN: 17410444. DOI: 10.1007/s11517-020-02260-3 (cit. a p. 6).
- [20] Laetitia Jeancolas, Graziella Mangone, Jean Christophe Corvol, Marie Vidailhet, Stéphane Lehericy, Badr Eddine Benkelfat, Habib Benali e Dijana Petrovska-Delacretaz. «Comparison of telephone recordings and professional microphone recordings for early detection of Parkinson's disease, using mel-frequency cepstral coefficients with Gaussian mixture models». In: *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH 2019-September* (2019). ISSN: 19909772. DOI: 10.21437/Interspeech.2019-2825 (cit. a p. 6).
- [21] David Gil e Magnus Johnsson. «Diagnosing Parkinson by using artificial neural networks and support vector machines». In: *Global Journal of Computer Science and Technology* 9 (4 2009). ISSN: 0975-4172 (cit. a p. 7).
- [22] Jan Ruzs, Roman Čmejla, Hana Růžičková, Jiří Klempíř, Veronika Majerová, Jana Picmausová, Jan Roth e Evžen Růžička. «Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test». In: *Movement Disorders* 26 (10 2011). ISSN: 08853185. DOI: 10.1002/mds.23680 (cit. a p. 7).
- [23] «Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings». In: *IEEE Journal of Biomedical and Health Informatics* 17 (4 2013). ISSN: 21682194. DOI: 10.1109/JBHI.2013.2245674 (cit. a p. 7).
- [24] Betül Erdogdu Sakar, Gorkem Serbes e C. Okan Sakar. «Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease». In: *PLoS ONE* 12 (8 2017). ISSN: 19326203. DOI: 10.1371/journal.pone.0182428 (cit. a p. 7).
- [25] Jan Mucha et al. «Identification of hypokinetic dysarthria using acoustic analysis of poem recitation». In: *2017 40th International Conference on Telecommunications and Signal Processing, TSP 2017* 2017-January (2017). DOI: 10.1109/TSP.2017.8076086 (cit. a p. 7).

- [26] Marius Ene. «Neural network-based approach to discriminate healthy people from those with Parkinson's disease». In: *Math. Comp. Sci. Ser* 35 (2008). ISSN: 1223-6934 (cit. a p. 7).
- [27] Frédéric Bimbot et al. «A tutorial on text-independent speaker verification». In: *Eurasip Journal on Applied Signal Processing* 2004 (4 2004). ISSN: 11108657. DOI: 10.1155/S1110865704310024 (cit. a p. 7).
- [28] T. Arias-Vergara, J. C. Vásquez-Correa e J. R. Orozco-Arroyave. «Parkinson's Disease and Aging: Analysis of Their Effect in Phonation and Articulation of Speech». In: *Cognitive Computation* 9 (6 2017). ISSN: 18669964. DOI: 10.1007/s12559-017-9497-x (cit. a p. 7).
- [29] Taoufiq Belhoussine Drissi, Soumaya Zayrit, Benayad Nsiri e Abdelkrim Ammoummou. «Diagnosis of Parkinson's disease based on wavelet transform and Mel Frequency Cepstral Coefficients». In: *International Journal of Advanced Computer Science and Applications* 10 (3 2019). ISSN: 21565570. DOI: 10.14569/IJACSA.2019.0100315 (cit. a p. 7).
- [30] Hao Fang, Chen Gong, Chen Zhang, Yanan Sui e Luming Li. «Parkinsonian Chinese Speech Analysis towards Automatic Classification of Parkinson's Disease». In: (2021). DOI: 10.48550/ARXIV.2105.14704. URL: <https://arxiv.org/abs/2105.14704> (cit. a p. 7).
- [31] Tamás Grósz, Róbert Busa-Fekete, Gábor Gosztolya e László Tóth. «Assessing the degree of nativeness and Parkinson's condition using Gaussian processes and deep rectifier neural networks». In: *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH 2015-January* (2015). ISSN: 19909772. DOI: 10.21437/interspeech.2015-195 (cit. a p. 7).
- [32] J. R. Orozco-Arroyave, J. C. Vásquez-Correa, F. Honig, J. D. Arias-Londono, J. F. Vargas-Bonilla, S. Skodda, J. Rusz e E. Noth. «Towards an automatic monitoring of the neurological state of Parkinson's patients from speech». In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* 2016-May (2016). ISSN: 15206149. DOI: 10.1109/ICASSP.2016.7472927 (cit. a p. 7).
- [33] Hanbin Zhang, Aosen Wang, Dongmei Li e Wenyao Xu. «DeepVoice: A voiceprint-based mobile health framework for Parkinson's disease identification». In: (2018), pp. 214–217. DOI: 10.1109/BHI.2018.8333407 (cit. a p. 7).
- [34] Evaldas Vaiciukynas, Antanas Verikas, Adas Gelzinis e Marija Bacauskiene. «Detecting Parkinson's disease from sustained phonation and speech signals». In: *PLoS ONE* 12 (10 2017). ISSN: 19326203. DOI: 10.1371/journal.pone.0185613 (cit. a p. 7).

- [35] Y. N. Zhang. «Can a Smartphone Diagnose Parkinson Disease? A Deep Neural Network Method and Telediagnosis System Implementation». In: *Parkinson's Disease 2017* (2017). ISSN: 20420080. DOI: 10.1155/2017/6209703 (cit. alle pp. 7, 21).
- [36] Athanasios Tsanas, Max A. Little, Patrick E. McSharry e Lorraine O. Ramig. «Accurate telemonitoring of parkinsons disease progression by noninvasive speech tests». In: *IEEE Transactions on Biomedical Engineering* 57 (4 2010). ISSN: 00189294. DOI: 10.1109/TBME.2009.2036000 (cit. a p. 7).
- [37] Juan Rafael Orozco-Arroyave et al. «Apkinson: The smartphone application for telemonitoring Parkinson's patients through speech, gait and hands movement». In: *Neurodegenerative Disease Management* 10 (3 2020). ISSN: 17582032. DOI: 10.2217/nmt-2019-0037 (cit. a p. 7).
- [38] María-José Marsano-Cornejo e Ángel Roco-Videla. «Comparison of the Acoustic Parameters Obtained With Different Smartphones and a Professional Microphone». In: *Acta Otorrinolaringologica (English Edition)* 73 (1 2022). ISSN: 21735735. DOI: 10.1016/j.otoeng.2020.08.009 (cit. alle pp. 8, 57).
- [39] Paul Boersma. «Praat, a system for doing phonetics by computer». In: *Glott International* 5 (9/10 2001) (cit. alle pp. 8, 14, 45).
- [40] Virgilijus Uloza, Evaldas Padervinskis, Aurelija Vegiene, Ruta Pribuisiene, Viktoras Saferis, Evaldas Vaiciukynas, Adas Gelzinis e Antanas Verikas. «Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameters and voice pathology screening». In: *European Archives of Oto-Rhino-Laryngology* 272 (11 2015). ISSN: 14344726. DOI: 10.1007/s00405-015-3708-4 (cit. alle pp. 8, 57).
- [41] Elizabeth U. Grillo, Jenna N. Brosious, Staci L. Sorrell e Supraja Anand. «Influence of Smartphones and Software on Acoustic Voice Measures.» In: *International Journal of Telerehabilitation* 8 (2 2016). ISSN: 1945-2020. DOI: 10.5195/ijtr.2016.6202 (cit. alle pp. 8, 57).
- [42] Jan Rusz, Tereza Tykalova, Lorraine O. Ramig e Elina Tripoliti. «Guidelines for Speech Recording and Acoustic Analyses in Dysarthrias of Movement Disorders». In: *Movement Disorders* 36 (4 2021). ISSN: 15318257. DOI: 10.1002/mds.28465 (cit. a p. 11).
- [43] Athanasios Tsanas, Max A. Little, Patrick E. McSharry e Lorraine O. Ramig. «Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity». In: *Journal of the Royal Society Interface* 8 (59 2011). ISSN: 17425662. DOI: 10.1098/rsif.2010.0456 (cit. a p. 12).

- 
- [44] João Paulo Teixeira, Carla Oliveira e Carla Lopes. «Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters». In: *Procedia Technology* 9 (2013). ISSN: 22120173. DOI: 10.1016/j.protcy.2013.12.124 (cit. alle pp. 12, 13).
- [45] Pedro Gómez-Vilda, Victoria Rodellar-Biarge, Víctor Nieto-Lluis, Cristina Muñoz-Mulas, Luis Miguel Mazaira-Fernández, Rafael Martínez-Olalla, Agustín Álvarez-Marquina, Carlos Ramírez-Calvo e Mario Fernández-Fernández. «Characterizing Neurological Disease from Voice Quality Biomechanical Analysis». In: *Cognitive Computation* 5 (4 2013). ISSN: 18669956. DOI: 10.1007/s12559-013-9207-2 (cit. a p. 12).
- [46] Athanasios Tsanas. «Accurate telemonitoring of Parkinson’s disease symptom severity using nonlinear speech signal processing and statistical machine learning». In: *Thesis* (June 2012) (cit. alle pp. 12, 13, 16, 19).
- [47] G. De Krom. «A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals». In: *Journal of Speech and Hearing Research* 36 (2 1993). ISSN: 00224685. DOI: 10.1044/jshr.3602.254 (cit. a p. 14).
- [48] Athanasios Tsanas, Max A. Little, Patrick E. McSharry, Jennifer Spielman e Lorraine O. Ramig. «Novel speech signal processing algorithms for high-accuracy classification of Parkinsons disease». In: *IEEE Transactions on Biomedical Engineering* 59 (2012) (cit. a p. 15).
- [49] Paul Boersma. «Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound». In: *Proceedings of the Institute of Phonetic Sciences* 17 (1993) (cit. a p. 15).
- [50] J. A. Gómez-García, L. Moro-Velázquez, J. D. Arias-Londoño e J. I. Godino-Llorente. «On the design of automatic voice condition analysis systems. Part III: review of acoustic modelling strategies». In: *Biomedical Signal Processing and Control* 66 (2021). ISSN: 17468108. DOI: 10.1016/j.bspc.2020.102049 (cit. a p. 15).
- [51] Lawrence R Rabiner e Ronald W Schafer. *Digital processing of speech signals*. Pearson Education India, 2010 (cit. a p. 15).
- [52] Florian Eyben, Felix Weninger, Florian Gross e Björn Schuller. «Recent developments in openSMILE, the munich open-source multimedia feature extractor». In: *MM 2013 - Proceedings of the 2013 ACM Multimedia Conference* (2013). DOI: 10.1145/2502081.2502224 (cit. alle pp. 15–17, 44).
- [53] Dmitry Bogdanov et al. «Essentia: An audio analysis library for music information retrieval». In: *Proceedings of the 14th International Society for Music Information Retrieval Conference, ISMIR 2013* (2013) (cit. alle pp. 15–18).



- [54] Francesco Piazza Holger Crysand Giovanni Tummarello. «Mpeg-7 Encoding and processing : MPEG7AUDIOENC+MPEG7AUDIODB». In: (2004) (cit. alle pp. 15, 16, 19).
- [55] Daniel McEnnis, Cory McKay e Ichiro Fujinaga. «jAudio: Additions and improvements». In: *ISMIR 2006 - 7th International Conference on Music Information Retrieval* (2006) (cit. alle pp. 15–17).
- [56] Benoit Mathieu, Slim Essid, Thomas Fillon, Jacques Prado e Gaël Richard. «Yaafe, an easy to use and efficient audio feature extraction software». In: *Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR 2010* (2010) (cit. alle pp. 16, 19).
- [57] Miron B. Kurşa e Witold R. Rudnicki. «Feature selection with the boruta package». In: *Journal of Statistical Software* 36 (11 2010). ISSN: 15487660. DOI: 10.18637/jss.v036.i11 (cit. a p. 22).
- [58] A Liaw e M Wiener. «Classification and Regression by randomForest». In: *R News* 2 (3 2002) (cit. a p. 22).
- [59] Federica Amato, Irene Rechichi, Luigi Borzì e Gabriella Olmo. «Sleep Quality through Vocal Analysis: a Telemedicine Application». In: (2022), pp. 706–711. DOI: 10.1109/PerComWorkshops53856.2022.9767372 (cit. a p. 22).
- [60] Chih Chung Chang e Chih Jen Lin. «LIBSVM: A Library for support vector machines». In: *ACM Transactions on Intelligent Systems and Technology* 2 (3 2011). ISSN: 21576904. DOI: 10.1145/1961189.1961199 (cit. alle pp. 22, 23).
- [61] Dr. R. Muthalagu, Mehak Musheer, Kumaresh Nandhini e N. Saranraj. «Type of Skin Disease Identification by Machine Learning Using Python». In: (2021) (cit. a p. 23).
- [62] Shichao Zhang, Xuelong Li, Ming Zong, Xiaofeng Zhu e Debo Cheng. «Learning k for kNN Classification». In: *ACM Transactions on Intelligent Systems and Technology* 8 (3 2017). ISSN: 21576912. DOI: 10.1145/2990508 (cit. a p. 23).
- [63] Daniel Berrar. *Bayes' theorem and naive bayes classifier*. Vol. 1-3. 2018. DOI: 10.1016/B978-0-12-809633-8.20473-1 (cit. a p. 24).
- [64] Alexey Natekin e Alois Knoll. «Gradient boosting machines, a tutorial». In: *Frontiers in Neurorobotics* 7 (DEC 2013). ISSN: 16625218. DOI: 10.3389/fnbot.2013.00021 (cit. a p. 24).
- [65] Tammy Harris e James W. Hardin. «Exact Wilcoxon signed-rank and Wilcoxon Mann-Whitney ranksum tests». In: *Stata Journal* 13 (2 2013). ISSN: 1536867X. DOI: 10.1177/1536867x1301300208 (cit. a p. 28).