**Avviso Pubblico “Riposizionamento competitivo RSR” - Regione Lazio Det. n. G18823**

Immagine che contiene Carattere, logo, Elementi grafici, simbolo

Descrizione generata automaticamente



**TELEMAIA**

**Deliverable D3.1.**

**< SW AI-based per il supporto alla diagnosi e alla valutazione del paziente>**

**Responsabili del Deliverable:**

Ingg. Michele Russo e Lorenzo Angeloni

- Sentech S.r.l.

e

Prof. Claudio Babiloni e Prof. Claudio Del Percio

- Sapienza Università di Roma

**SW AI-BASED PER IL SUPPORTO ALLA DIAGNOSI E ALLA VALUTAZIONE DEL PAZIENTE**

**Sommario**

**Introduzione 4**

***Scopo dello studio TELEMAIA e del Deliverable 3.1* 6**

**Metodi 7**

***Partecipanti e criteri diagnostici* 7**

***Il telemonitoraggio domestico* 10**

***Studio dell’accuratezza discriminativa tra i gruppi Nold e PDCD dei marker di rsEEG, MRI e di telemonitoraggio domestico* 11**

***Le registrazioni e le analisi dei dati rsEEG* 13**

***Registrazione delle immagini di risonanza magnetica (MRI)* 13**

***Marker MRI (features)* 15**

***Strumenti di intelligenza artificiale (AI) per machine learning* 17**

***Fase 1 dell'analisi statistica: selezione dei marker di rsEEG, MRI e di telemonitoraggio domestico per il machine learning* 21**

***Fase 2 dell'analisi statistica: associazione mediante machine learning tra i marker di funzione cognitiva globale (MMSE) e i marcatori (caratteristiche) di rsEEG, MRI e di telemonitoraggio domestico in partecipanti Nold e pazienti con declino cognitivo lieve (MCI) dovuto a malattia neurodegenerativa (AD e PD)* 25**

**Risultati 25**

***Analisi dei marker rsEEG* 25**

***Analisi dei marker MRI* 28**

***Analisi dei marker di telemonitoraggio domestico delle funzioni cognitive (app SmartMe&You)* 31**

***Conclusioni* 35**

***Bibliografia* 38**

**Appendice sulle tecniche di machine learning 40**

**Introduzione**

La malattia di Parkinson (PD) è caratterizzate da una progressiva neurodegenerazione dei neuroni encefalici dovuta ai corpi di Lewy intracellulari, esito di accumulo patologico dell'alfa-sinucleina (<https://www.epicentro.iss.it/parkinson/>). Il processo neuropatologico si sviluppa inizialmente nei neuroni dopaminergici della sostanza nera del mesencefalo con disfunzioni dei gangli della base e dei sistemi motori.

Il PD è il secondo disturbo neurodegenerativo progressivo più comune dopo il morbo di Alzheimer (AD), con una prevalenza stimata di oltre 8,5 milioni di casi a livello globale, con un maggior numero di casi nei maschi rispetto alle femmine (Organizzazione Mondiale della Sanità, 2019). Secondo le linee guida cliniche attuali, i pazienti con PD presentano principalmente sintomi motori come acinesia o bradicinesia, tremore involontario degli arti o altre parti del corpo, instabilità posturale e del passo e rigidità musco-scheletrica (Aarsland et al., 2021). Tuttavia, durante il decorso della malattia, spesso si osservano anche deficit cognitivi nei pazienti con PD, che includono difficoltà nella pianificazione, nella memoria di lavoro, nelle capacità esecutive frontali, nell’attenzione, nella fluidità verbale semantica e nelle abilità visuo-spaziali (Levy et al., 2002; Aarsland et al., 2003; Buter et al., 2008; Fang et al., 2020).

Nei pazienti con PD, i deficit cognitivi diventano più marcati in relazione all’età. Circa il 50% dei pazienti con PD sviluppa deficit cognitivi (PDCD) e demenza entro 10 anni dalla diagnosi, mentre la maggior parte dei pazienti con PD sviluppa demenza entro 20 anni dalla diagnosi (Aarsland et al., 2021). Inoltre, circa il 50% dei pazienti con PD presenta un lieve deterioramento cognitivo (MCI) al momento della diagnosi, che aumenta del 40-50% dopo 5 anni di follow-up (Aarsland et al., 2010, 2021; Pedersen et al., 2017; Lawson et al., 2016). Infine, si osservano disfunzioni anche di carattere non-cognitivo e non-motorio, quali i disturbi della vigilanza e le allucinazioni visive e i disturbi del sonno (Gunn et al., 2014, PMID: 24411329; Scullin et al., 2015, PMID: 25588355).

Nei pazienti PD con deficit cognitivo (PDCD) e i disturbi della vigilanza sono correlati ad anomalie funzionali nelle reti neurali frontali, temporali e parietali della corteccia cerebrale, come indicato dalle registrazioni elettroencefalografiche in veglia rilassata (rsEEG; Peláez Suárez et al., 2021) e di risonanza magnetica (MRI; Lin et al., 2021). In questo contesto, i gruppi di ricerca delle due Unità Operative cliniche del Progetto TELEMAIA, la Sapienza Università di Roma e l’IRCCS San Raffaele Pisana S.r.l., hanno dato contributi scientifici significativi come riportato sulle pagine WEB del Consorzio internazionale PDWAVES ([www.pdwaves.eu](http://www.pdwaves.eu/)) e su vari articoli scientifici (Babiloni et al., 2020; Babiloni et al., 2019; Babiloni et al., 2018a; Babiloni et al., 2018b; Babiloni et al., 2017). In questi articoli, i pazienti con PDCD hanno mostrato un aumento anormale di ritmi EEG alla frequenza del ritmo delta (< 4 Hz) su vaste regioni dello scalpo (Babiloni et al., 2020; Babiloni et al., 2019; Babiloni et al., 2018a; Babiloni et al., 2018b; Babiloni et al., 2017), modulate in ampiezza dalla somministrazione del farmaco di elezione per i pazienti con PD, la levodopa (Babiloni et al., 2019).

I fattori di rischio di malattia endogeni (genotipizzazione, fragilità, sesso, età) ed esogeni (stile di vita) e le tecniche di intelligenza artificiale (AI) possono essere usati in modo fruttuoso (1) per riconoscere e inquadrare clinicamente i pazienti con PDCD e AD sulla base dei biomarcatori (marker) MRI e rsEEG e (2) per guidare strategie di prevenzione e intervento (Teipel et al, 2022; Babiloni et al., 2021-2020; Zarifkar et al., 2021; Fiorenzato et al., 2019; Kaasinen et al., 2021; Picillo et al., 2021; Colombo et al., 2015).

Le tecniche AI di machine learning utilizzano algoritmi complessi supervisionati dall’esperto per analizzare grandi insiemi di dati, identificando modelli di relazione tra le variabili usate come input per le classificazioni o le previsioni cliniche che potrebbero non essere discernibili attraverso metodi statistici tradizionali, quali il calcolo dell’area sotto la curva ROC, la regressione logistica, etc. (<https://it.wikipedia.org/wiki/Apprendimento_automatico>). Tra le tecniche di machine learning più consolidate, si annovano l’algoritmo k-nearest neighbors (K-NN), il support vector machine (SVM), il random forest (RF), la discriminant analysis (DA), etc.

Studi precedenti hanno usato tecniche di machine learning per analizzare le registrazioni vocali di pazienti con PD e il tremore involontario, rilevabili in principio anche mediante telemonitoraggio domestico, hanno raggiunto un tasso di accuratezza superiore all'85% nel distinguere quei pazienti da individui di controllo Nold (Hossain e Amenta, 2024; Sajal et al., 2020).

Altri studi hanno usato tecniche SVM di machine learning, ottenendo un'elevata precisione (> 80%) nella predizione del declino cognitivo lieve (MCI) in pazienti con PD sulla base di marker di MRI strutturale (Beheshti and Ko, 2024). Altri studi hanno mostrato che tecniche SVM di machine learning forniscono classificazioni accurate > 80% tra pazienti PD e partecipanti di controllo sulla base di marker rsEEG e MRI, in alcuni casi combinati (Chaturvedi et al., 2019; Zhang et al., 2020, 2022).

*Scopo dello studio TELEMAIA e del Deliverable 3.1.*

Lo studio clinico del progetto TELEMAIA ha l'ipotesi primaria di verificare se gli strumenti di machine learning possano prevedere con precisione i marker derivati dalla raccolta di dati effettuata in pazienti con PDCD in ambito ospedaliero (ad esempio, profilo neuropsicologico e clinico, rsEEG, e MRI), sulla base di marker demografici (i.e., sesso, età, livello di istruzione) e marker di telemonitoraggio domestico derivati dalla raccolta di dati per una settimana con dispositivi commerciali, come smartwatch (attività fisica, ciclo sonno-veglia, variabilità della frequenza cardiaca, qualità del sonno notturno), tablet (videogiochi seri che misurano la fluttuazione delle prestazioni cognitive e profilatori del sonno che forniscono un ipnogramma e indicatori di qualità del sonno durante il riposo notturno). Se confermati, i marker demografici e quelli di telemonitoraggio domestico potrebbero surrogare i costosi marker clinici e strumentali derivati da ambienti ospedalieri, con vantaggio per la spesa pubblica e per i pazienti con PDCD e le loro famiglie.

In questo Deliverable 3.1. del Progetto TELEMAIA, descriviamo le procedure di machine learning usate per due scopi:

(1) La classificazione tra partecipanti Nold e pazienti con declino cognitivo lieve (MCI) dovuto a malattia neurodegenerativa (AD e PD) sulla base dei marker di rsEEG, MRI e telemonitoraggio domestico delle funzioni cognitive con l’app “SmartMe&You”. Essa è costituita da una batteria di 7 test neuropsicologici eseguibili su tablet commerciali per la valutazione delle funzioni cognitive, con un focus sulle funzioni esecutive frontali e l’attenzione visuo-spaziale. La validazione dell’app SmartMe&You è stata effettuata in pazienti con PD e con malattia di Alzheimer nel corso di un progetto finanziato dal Ministero della Salute Italiano: il progetto di Ricerca Finalizzata dal titolo breve "GRIDCORE", completato alla fine del 2016 (<https://www.garr.it/en/news-events/10-focus-on/649-gridcore-strumenti-innovativi-contro-le-malattie-neurodegenerative>).

(2) Lo studio di associazione tra i suddetti marker di telemonitoraggio domestico delle funzioni cognitive (arricchiti con i marker demografici) e i marker di rsEEG delle funzione cerebrale e i punteggi di mini-mental state examination (MMSE) che riflettono lo stato cognitivo globale. La conferma dell’ipotesi incoraggerebbe l’uso del telemonitoraggio domestico delle funzioni cognitive per i controlli periodici nel corso di una condizione di PDCD che progredisce nel corso di anni nei pazienti con PD e richiede numerose visite in ospedale.

**Metodi**

*Partecipanti e criteri diagnostici*

I dataset utilizzati per le analisi di machine learning descritte in questo Deliverable 3.1. sono costituiti da dati clinici, neuropsicologici, MRI, rsEEG e di telemonitoraggio domestico ottenuti in pazienti con declino cognitivo lieve (MCI) dovuto a malattia neurodegenerativa (AD e PD) e anziani Nold in parte raccolti nel corso del Progetto TELEMAIA e in parte disponibili negli archivi delle due Unità Operative cliniche dell’IRCCS San Raffaele Roma S.r.l. e Sapienza Università di Roma. In questo modo, è stato possibile avere un numero di dataset tale da rendere più validi e replicabili i risultati delle analisi di machine learning. Tutti i dataset usati in questo Deliverable 3.1 derivano da attività cliniche svolte con l'approvazione istituzionale e il consenso esplicito di ogni partecipante, in linea con il Codice Etico dell'Associazione Medica Mondiale (Dichiarazione di Helsinki) e con gli standard stabiliti dall'Institutional Review Board locale.

In linea con i precedenti studi del progetto PDWAVES delle menzionate Unità Operative cliniche (Babiloni et al., 2020; Babiloni et al., 2019; Babiloni et al., 2018a; Babiloni et al., 2018b; Babiloni et al, 2017), la diagnosi di PD si basa su una misura standard di tremore, rigidità e bradicinesia secondo le linee guida della United Kingdom PD Society Brain Bank (Hughes et al., 1992) e i “Criteri di diagnosi della malattia di Parkinson” (Gelb et al., 1999).

Criteri di inclusione per i pazienti con PDCD

● Età > 60 anni.

● Piena comprensione della lingua italiana.

● Disponibilità/abilità a partecipare alla parte di monitoraggio domestico del progetto (per i pazienti con PDCD la presenza di un familiare in casa che possa dare una significativa assistenza per le procedure di telemonitoraggio domestico e per procedure di stimolazione transcranica a correnti di bassa intensità).

● Capacità di firmare il consenso informato.

Diagnosi di PD basata su una misura standard di tremore, rigidità e bradicinesia, secondo le linee guida del United Kingdom PD Society Brain Bank (Hughes et al., 1992) e dei “Criteria of diagnosis of Parkinson disease (Gelb et al., 1999).

Criteri di esclusione dei pazienti con PDCD e dei partecipanti Nold

● Condizioni che a giudizio dello sperimentatore possano interferire con lo studio, oppure ne sconsiglino la partecipazione per motivi di sicurezza.

● Uso di pacemaker, controindicati per le registrazioni di MRI e per l’uso di smartwatch.

● Uso di pacemaker, controindicati per le registrazioni di MRI e per l’uso di smartwatch.

● Partecipanti a cui sia stato somministrato un farmaco sperimentale nei 30 giorni precedenti l’inizio dello studio.

● Partecipanti che non abbiano firmato il consenso informato allo studio.

● Le seguenti forme di Parkinsonismo: (i) malattia di Lewy Body; (ii) parkinsonismo indotto da farmaci; (iii) parkinsonismo cerebrovascolare; (iv) parkinsonismo atipico con risposte assenti o minime ai farmaci antiparkinsoniani.

● Presenza di grandi patologie (a parte il PD nei pazienti con PDCD) che possano condizionare la diagnosi e/o l’esito delle prove.

● Ogni tipo di demenza.

● Disturbi mentali maggiori, dovuti a cause neurologiche e psichiatriche.

● Partecipanti a cui sia stato somministrato un farmaco sperimentale nei 30 giorni precedenti l’inizio dello studio.

● Partecipanti che non abbiano firmato il consenso informato allo studio.

● Le seguenti forme di Parkinsonismo: (i) malattia di Lewy Body; (ii) parkinsonismo indotto da farmaci; (iii) parkinsonismo cerebrovascolare; (iv) parkinsonismo atipico con risposte assenti o minime ai farmaci antiparkinsoniani.

● Presenza di grandi patologie (a parte il PD nei pazienti con PDCD) che possano condizionare la diagnosi e/o l’esito delle prove.

● Ogni tipo di demenza.

● Disturbi mentali maggiori, dovuti a cause neurologiche e psichiatriche.

Riguardo lo stato cognitivo globale e il grado di autonomia nello svolgimento delle attività comuni della vita quotidiana (ad esempio, lavarsi, vestirsi, prepararsi un pasto, fare acquisti, etc.), si considerino i seguenti indicatori clinici:

● Indicatori di screening per POSSIBILE DEMENZA (la diagnosi è solo clinica)

The Montreal Cognitive Assessment (MoCA) ≤ 17

Mini-Mental State Examination (MMSE) ≤ 18

Parkinson's Disease - Cognitive Functional Rating Scale (PD-CFRS) PD-CFRS > 6.5

Activities of daily living (ADL) e Instrumental ALD (IADL) riportanti almeno un deficit per ciascuno

● Indicatori di screening per POSSIBILE MCI sono:

MOCA < 23

MMSE < 25

ADL = 6

IADL (può essere alterata) < 1

PD-CFRS ≤ 6.5

Per lo studio delle singole funzioni cognitive, viene usata una batteria standard per lo studio di persone anziane con deficit cognitivi con test specifici per la memoria episodica verbale e visuo-spaziale, la memoria a breve termine di numeri, le funzioni esecutive frontali, la fluidità verbale per fonemi e per categorie (ad esempio, animali e piante), le abilità visuo-spaziali di copia di modelli, etc.

*Il telemonitoraggio domestico*

Per il telemonitoraggio domestico, viene usata l’attuale piattaforma “SmartMe&You” basata sul WEB, operativa sulla rete cloud del Consorzio GARR ([http://smartme.cloud.garr.it](http://smartme.cloud.garr.it/)). Essa consente la raccolta in tempo reale, la gestione e la visualizzazione dei dati del telemonitoraggio domestico delle funzioni cognitive e motorie, di stress cardiovascolare generale e del ciclo sonno-veglia e della qualità del sonno notturno negli utilizzatori dei dispositivi di telemonitoraggio.

La valutazione delle funzioni cognitive durante il telemonitoraggio domestico è effettuata con l’app “SmartMe&You”, mentre la valutazione della motricità quotidiana, del ritmo sonno-veglia e della variabilità del battito cardiaco come indice dello stress cardiovascolare generale è valutata attraverso i dati raccolti dallo smartwatch commerciale Samsung Galaxy Watch4 44mm (<https://www.samsung.com/it/watches/galaxy-watch/galaxy-watch4-green-bluetooth-sm-r870nzgaitv/buy/>).

*Studio dell’accuratezza discriminativa tra i gruppi Nold e PDCD dei marker di rsEEG, MRI e di telemonitoraggio domestico*

Il numero e le caratteristiche delle persone con PDCD e Nold considerate in questo studio esplorativo sui marker rsEEG sono riportati nella **Tabella 1**.

**Tabella 1**

|  |  |  |  |
| --- | --- | --- | --- |
| **Informazioni demografiche e cliniche nei gruppi Nold e PDCD (rsEEG)** | | | |
|  | **Nold** | **PDCD** | **Confronti statistici** |
| **N** | 54 | 50 | -- |
| **Età (anni)** | 69.1 ± 1.1 SE | 70.5 ± 0.7 SE | T test: n.s. |
| **Genere (M/F)** | 25/29 (46%) | 39/11 (78%) | Fisher test: n.s. |
| **Istruzione (anni)** | 10.5 ± 0.6 SE | 9.5 ± 0.4 SE | T test: n.s. |
| **MMSE** | 28.5 ± 0.2 SE | 20.3 ± 0.4 SE | Mann Whitney U test: n.s. |

**Tabella 1**. *Informazioni demografiche e cliniche rilevanti dei pazienti con malattia di Parkinson con deficit cognitivo (PDCD) e degli anziani normali non compromessi cognitivamente (Nold). La tabella riporta anche i risultati delle analisi statistiche (p < 0,05 non corretto) calcolate per valutare la presenza o l'assenza di differenze statisticamente significative dei marker rsEEG tra i gruppi Nold e PDCD per quanto riguarda l'età (t-test), il sesso (maschio/femmina; test esatto di Fisher), il livello di istruzione (t-test) e il punteggio della mini-mental state evaluation (MMSE) (test U di Mann-Whitney).*

Il numero e le caratteristiche delle persone con PDCD (MCI) e Nold considerate in questo studio esplorativo sui marker MRI sono riportati nella Tabella 2.

**Tabella 2**

|  |  |  |  |
| --- | --- | --- | --- |
| **Informazioni demografiche e cliniche nei gruppi Nold e PDCD (MRI)** | | | |
|  | **Nold** | **PDCD** | **Confronti statistici** |
| **N** | 32 | 43 | -- |
| **Età (anni)** | 65.5 ± 1.5 SE | 66.0 ± 1.4 SE | T test: n.s. |
| **Genere (M/F)** | 10/22 (31%) | 31/11 (72%) | Fisher test: p < 0.05 |
| **Istruzione (anni)** | 9.3 ± 0.2 SE | 6.9 ± 0.8 SE | T test: p < 0.05 |
| **MMSE** | 27.9 ± 0.1 SE | 23.4 ± 0.6 SE | Mann Whitney U test: p < 0.05 |

**Tabella 2**. *Informazioni demografiche e cliniche rilevanti dei partecipanti con PDCD e quelli Nold. La tabella riporta anche i risultati delle analisi statistiche (p < 0,05 non corretto) calcolate per valutare la presenza o l'assenza di differenze statisticamente significative dei marker MRI tra i gruppi PDCD e Nold per quanto riguarda l'età (t-test), il sesso (maschio/femmina; test esatto di Fisher), il livello di istruzione (t-test) e il punteggio della mini-mental state evaluation (MMSE) (test U di Mann-Whitney).*

Il numero e le caratteristiche delle persone con declino cognitivo lieve (MCI) dovuto a malattia neurodegenerativa (AD e PD) e quelle Nold considerate in questo studio esplorativo sui marker di telemonitoraggio domestico sono riportati nella Tabella 3.

**Tabella 3**

|  |  |  |  |
| --- | --- | --- | --- |
| **Informazioni demografiche e cliniche nei gruppi Nold e MCI**  **(Telemonitoraggio domestico)** | | | |
|  | **Nold** | **MCI** | **Confronti statistici** |
| **N** | 27 | 20 | - |
| **Età (anni)** | 53.8 ± 4.0 SE | 77.3 ± 1.2 SE | T-test: p < 0.0001 |
| **Genere (M/F)** | 15/12 (56%) | 16/4 (78%) | Fisher test: n.s. |
| **Istruzione (anni)** | 16.4 ± 0.4 SE | 12.0 ± 0.9 SE | T-test: p < 0.0001 |
| **MMSE** | 29.6 ± 0.2 SE | 25.9 ± 0.7 SE | Mann Whitney U test: p < 0.0001 |

**Tabella 3**. *Informazioni demografiche e cliniche rilevanti dei partecipanti con deficit cognitivo lieve (MCI) dovuto a malattia neurodegenerativa (PD e AD) e quelli non compromessi cognitivamente (Nold). La tabella riporta anche i risultati delle analisi statistiche (p < 0,05 non corretto) calcolate per valutare la presenza o l'assenza di differenze statisticamente significative dei marker di telemonitoraggio domestico tra i gruppi Nold e MCI per quanto riguarda l'età (t-test), il sesso (maschio/femmina; test esatto di Fisher), il livello di istruzione (t-test) e il punteggio della mini-mental state evaluation (MMSE) (test U di Mann-Whitney).*

*Le registrazioni e le analisi dei dati rsEEG*

Per la registrazione e l’analisi quantitativa dei dati rsEEG, si veda il Deliverable 2.2. (“Estrazione dei marcatori EEG e relativa analisi statistica”).

*Registrazione delle immagini di risonanza magnetica (MRI)*

Tutte le scansioni MRI sono state eseguite nei partecipanti Nold e PDCD utilizzando scanner da 3,0 Tesla attivi presso l'IRCCS San Raffaele Roma e la Sapienza Università di Roma. Ciascuna marca di scanner è utilizzata nelle persone Nold e nei pazienti PDCD ed è considerata un confondente di banco nell'analisi statistica di gruppo per la caratterizzazione della riproducibilità.

Il protocollo di MRI comprende scansioni anatomiche T1 di interesse per questo studio. I dati di MRI acquisiti sono anonimizzati secondo gli standard internazionali per proteggere i dati biomedici sensibili. Il gruppo di lavoro dei partecipanti della Sentech s.r.l. ha eseguito l'analisi a livello centrale. Prima dell'analisi, tutti i dati sono stati sottoposti a un'ispezione visiva per verificare la qualità (ad esempio, artefatti visibili, tra cui il movimento della testa, l'avvolgimento, le interferenze a radiofrequenza e le disomogeneità dell'intensità del segnale o del contrasto). I dati di MRI sono formattati secondo gli standard internazionali Brain Imaging Data Structure (BIDS).

I risultati inclusi in questo progetto provengono da una pre-elaborazione eseguita con fMRIPrep 20.2.0 (Esteban, et al., 2018; 2019), che si basa su Nipype 1.5.1 (Gorgolewski et al., 2011). Le immagini pesate in T1 (T1w) si trovano nel set di dati BIDS di input. Ogni immagine T1w viene corretta per la non uniformità dell'intensità (INU) con N4BiasFieldCorrection (Tustison et al., 2010), distribuito con ANTs 2.3.3 (Avants et al., 2008), e utilizzato come riferimento T1w in tutto il flusso di lavoro. Il riferimento T1w viene quindi sottoposto a skull-stripping con un'implementazione Nipype del flusso di lavoro antsBrainExtraction.sh (da ANTs), utilizzando OASIS30ANTs come template di destinazione. La segmentazione del tessuto cerebrale del liquido cerebrospinale (CSF), della materia bianca (WM) e della materia grigia (GM) viene eseguita sul T1w estratto dal cervello utilizzando fast (FSL 5.0.9, Zhang, et al., 2001). Le superfici cerebrali sono ricostruite con recon-all (FreeSurfer 6.0.1, Dale, et al., 1999) e la maschera cerebrale stimata in precedenza è raffinata con una variazione personalizzata del metodo per riconciliare le segmentazioni del GM corticale di Mindboggle derivate da ANTs e FreeSurfer (Klein et al. 2017). La normalizzazione spaziale basata sul volume a uno spazio standard (MNI152NLin2009cAsym) viene eseguita mediante registrazione non lineare con antsRegistration (ANTs 2.3.3), utilizzando versioni estratte dal cervello sia del riferimento T1w che del template T1w. Per la normalizzazione spaziale è stato selezionato il seguente template: ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009; TemplateFlow ID: MNI152NLin2009cAsym). Molte operazioni interne di fMRIPrep utilizzano Nilearn 0.6.2 (Abraham et al., 2014), soprattutto all'interno del flusso di lavoro dell'elaborazione funzionale. Per ulteriori dettagli sulla pipeline, consultare la sezione sul flusso di lavoro nella documentazione di fMRIPrep (<https://fmriprep.org/en/20.2.0/workflows.html>).

*Marker MRI (features)*

Si sono estratti dai dati di MRI 7 network corticali messi in luce da studi di MRI funzionale registrata in condizione di veglia quieta nel corso degli ultimi 10 anni e rappresentati nell'atlante di Yeo, usato come riferimento per il Progetto TELEMAIA (Yeo et al., 2011).

Per l’analisi dei dati di MRI, abbiamo usato la piattaforma freeware di analisi “FreeSurfer”. Essa fornisce lo spessore corticale medio, l'area superficiale e il volume di ogni rete e di ogni partecipante. Le nostre analisi sono concentrate sui marcatori di MRI di seguito indicati (Yeo, 2011).

DEFAULT MODE NETWORK (DMN)

● volume

● cortical thickness

FRONTOPARIETAL NETWORK (FPN)

● volume

● cortical thickness

VISUAL NETWORK (VN)

● volume

● cortical thickness

SOMATOMOTOR NETWORK (SM)

● volume

● cortical thickness

DORSAL ATTENTION NETWORK (DAN)

● volume

● cortical thickness

VENTRAL ATTENTION NETWORK (VAN)

● volume

● cortical thickness

LIMBIC NETWORK (LN)

● volume

● cortical thickness

MESIAL TEMPORAL NETWORK (MTN)

● volume hippocampus

● volume amygdala

BASAL GANGLIA AND MIDBRAIN NETWORK (BGMN)

● volume n. caudate

● volume n. putamen

● volume n. pallidum

● volume n. accumbens.

● thalamus.

GLOBAL MARKERS

● global white matter (WM) volume

● global grey matter (GM) volume

● global cerebrospinal fluid (CSF) volume.

*Strumenti di intelligenza artificiale (AI) per machine learning*

Come accennato in precedenza, utilizzeremo strumenti AI indipendenti di machine learning per classificare i partecipanti con deficit cognitivo (PDCD ed MCI) e quelli Nold asulla base dei dati raccolti in ambito ospedaliero (rsEEG, MRI) e dei marker di telemonitoraggio domestico delle funzioni cognitive mediante le prestazioni (accuratezza %) dei 7 video-giochi seri dell’app “SmartMe&You”, considerando anche i marker demografici (i.e., sesso, età, livello di istruzione). Inoltre verrà studiata l’associazione tra i suddetti marker di telemonitoraggio domestico delle funzioni cognitive (arricchiti con i marker demografici) e i marker di rsEEG delle funzione cerebrale e i punteggi di mini-mental state examination (MMSE) che riflettono lo stato cognitivo globale.

In particolare, gli strumenti di AI saranno i seguenti: Analisi discriminante (DA), Generalized Linear Regression Model (GLRM), K-nearest neighbor (k-NN), Random Forest (RF) e Support Vector Machine (SVM). Per la classificazione, le prestazioni in uscita di questi strumenti di AI vengono misurate in termini di accuratezza %, sensibilità %, specificità %, precisione %, ed F1.score %. Per ogni classificatore. Per quanto riguarda lo studio di associazione, le prestazioni in uscita di questi strumenti di AI vengono misurate in termini di Mean Squared Error (ME), Root Mean Squared Error (RMSE) e Mean Absolute Error (MAE).

*Analisi discriminante (DA)*

DA è un approccio utilizzato nell'apprendimento automatico supervisionato per risolvere problemi di classificazione multi-classe. DA funziona identificando una combinazione lineare o non-lineare di caratteristiche che separa o caratterizza due o più classi di oggetti o eventi. A tal fine, la DA proietta i dati a due o più dimensioni in un'unica dimensione, in modo che possano essere classificati più facilmente. Per questo motivo, la tecnica viene talvolta definita riduzione della dimensionalità. Questa tecnica è importante nella scienza dei dati perché aiuta a ottimizzare i modelli di apprendimento automatico. Segue un modello generativo. Ciò significa che gli algoritmi DA modellano la distribuzione dei dati per ciascuna classe e utilizzano il teorema di Bayes per classificare i nuovi punti di dati. Questo approccio si è dimostrato proficuo per la selezione congiunta di caratteristiche e campioni e per una valutazione di pazienti con PD a partire da dati di MRI (Adeli et al., 2016).

*Generalized Linear Regression Model (GLRM)*

I modelli GLM includono ed estendono la classe dei modelli lineari. Essi sono più permissivi riguardo le restrizioni sui modelli lineari, che spesso nella pratica vengono violate. Ad esempio, le risposte binarie (sì/no o 0/1) non presentano la stessa varianza tra le classi. Il GLM è una tecnica di modellazione parametrica. I modelli parametrici fanno ipotesi sulla distribuzione dei dati. Quando le ipotesi sono soddisfatte, i modelli parametrici possono essere più efficienti dei modelli non parametrici. La sfida nello sviluppo di modelli di questo tipo consiste nel valutare la misura in cui le ipotesi vengono soddisfatte. Per questo motivo, la diagnostica di qualità è fondamentale per sviluppare modelli parametrici di qualità.

GLM può essere utilizzato per creare modelli di classificazione o regressione:

* Classificazione: la regressione logistica binaria è l'algoritmo di classificazione GLM. L'algoritmo utilizza la funzione di collegamento logit e la funzione di varianza binomiale.
* Regressione: la regressione lineare è l'algoritmo di regressione GLM. L'algoritmo non presuppone alcuna trasformazione del target e una varianza costante nell'intervallo dei valori target.

*Random Forest (RF)*

Random Forest è un algoritmo di machine learning comunemente utilizzato, che combina l'output di più strutture ad albero decisionali per raggiungere un unico risultato. La facilità d'uso e la flessibilità ne hanno favorito l'adozione, in quanto gestisce sia i problemi di classificazione che quelli di regressione.

Le strutture ad albero decisionali iniziano con una domanda di base, associate ad una serie di domande per determinare una risposta. Queste domande formano i nodi decisionali nella struttura e rappresentano un mezzo per suddividere i dati. Ogni domanda concorre alla decisione finale, che viene indicata dal nodo foglia. Le osservazioni che soddisfano i criteri seguiranno il ramo "Sì" e quelle che non lo fanno seguiranno il percorso alternativo. Le strutture ad albero decisionali cercano di trovare il modo migliore per suddividere i dati e sono in genere addestrate attraverso l'algoritmo CART (Classification and Regression Tree). È possibile utilizzare metriche, come l'impurità di Gini, il guadagno di informazioni o l'errore quadratico medio (mean square error, MSE) per valutare la qualità della suddivisione.

Nonostante le strutture ad albero decisionali siano algoritmi di apprendimento supervisionato comuni, possono essere soggette a problemi, quali distorsione (bias) e overfitting. Tuttavia, quando più strutture ad albero decisionali formano un insieme nell'algoritmo Random Forest, prevedono risultati più accurati, in particolare quando le singole strutture ad albero non sono correlate tra loro.

*K-Nearest Neighbour (k-NN)*

L'algoritmo K-nearest neighbor (k-NN) è un noto metodo non parametrico che può essere utilizzato per la classificazione e la regressione (Bishop, 2006). Secondo questo approccio, un campione sconosciuto viene etichettato con l'etichetta più comune tra i suoi k vicini più prossimi nell'insieme di addestramento. La logica alla base del classificatore 3-NN è che, dato un campione sconosciuto 𝐱 da assegnare a una delle 𝑐𝑖 classi del problema in questione, le probabilità a-posteriori 𝑝(𝑐𝑖 |𝐱) di 𝐱 possono essere stimate come segue:

𝑝(𝑐𝑖 |𝐱) = 𝑛𝑖∕K

dove 𝑛𝑖 è il numero di campioni tra i 𝐾 vicini di 𝐱 appartenenti alla 𝑖esima classe.

*Support Vector Machine (SVM)*

Il Support Vector Machine (SVM) è un algoritmo di machine learning supervisionato con max-margin utilizzato per attività di classificazione, regressione e persino rilevamento di valori anomali lineari o non lineari. L’obiettivo di un algoritmo SVM è quello di trovare un iperpiano che separi, al miglior grado possibile, i punti di dati di una classe da quelli di un’altra classe. Per “migliore” si intende l’iperpiano che ha il margine maggiore tra le due classi, rappresentate dal più e dal meno nella figura qui sotto. Per margine si intende la larghezza massima della linea parallela all’iperpiano che non ha punti di dati interni. L’algoritmo è in grado di trovare un iperpiano di questo tipo solo per i problemi separabili in modo lineare, mentre per i problemi più pratici l’algoritmo massimizza il margine soft, il che consente un numero ridotto di classificazioni errate. Il Support Vector Machine rientra in una classe di algoritmi di Machine Learning definiti metodi kernel, in cui è possibile trasformare le feature per mezzo di una funzione kernel. Le funzioni kernel mappano i dati su uno spazio diverso e spesso di dimensione superiore con l’idea che, dopo questa trasformazione, le classi saranno più facili da separare, semplificando potenzialmente i confini decisionali non lineari complessi in confini lineari nello spazio di dimensione superiore delle feature mappate. In questo processo, non è necessario trasformare esplicitamente i dati, il che sarebbe molto dispendioso a livello computazionale. Questo è tradizionalmente noto come il “kernel trick”.

*Fase 1 dell'analisi statistica: selezione dei marker di rsEEG, MRI e di telemonitoraggio domestico per il machine learning*

La prima sessione statistica mira a selezionare le caratteristiche di rsEEG, MRI e di telemonitoraggio domestico che mostrano differenze statisticamente significative tra i gruppi Nold e PDCD. A tal fine, viene eseguita come esposto nelle sezioni che seguono.

Selezione dei marker mediante modelli ANOVA: differenze nei marcatori rsEEG, MRI e telemonitoraggio domestico (features) tra i gruppi Nold e con deficit cognitivo (PDCD ed MCI)

Nella prima fase della procedura, calcoliamo modelli ANOVA (p < 0,05) con il fattore Gruppo (Nold e PDCD) e le caratteristiche rsEEG, MRI e di telemonitoraggio domestico come variabili dipendenti. L'Unità Clinica (sede di registrazione presso l'IRCCS San Raffaele Roma e Sapienza Università di Roma) è utilizzata come covariata nell'analisi.

I fattori ANOVA per le variabili rsEEG sono Gruppo (Nold e PDCD), Regione di Interesse - ROI - (frontale, centrale, parietale, temporale, occipitale e limbico) e Banda (delta, theta, alfa 2, alfa 3, beta 1, beta 2 e gamma).

Per i marker MRI, usiamo il modello ANOVA sui marker MRI dei network corticali, con i fattori Gruppo (Nold e PDCD) e Network corticale (DMN, FPN, VN, SN, DAN, VAN, LN) e il volume di materia grigia normalizzata al volume cranico totale (TIV) come variabile dipendente.

I fattori ANOVA per il telemonitoraggio domestico delle funzioni cognitive sono Gruppo (Nold e PDCD) e Videogioco (7 videogiochi seri su tablet implementi nell'app “SmartMe&You”, ogni videogioco serio come livello del modello ANOVA) e la percentuale complessiva di risposte corrette è la variabile dipendente. Nel Deliverable 3.1., le analisi dei marker sono state condotte solo sulle variabili di prestazione dei 7 videogiochi seri, poiché l’estrazione dei marker delle altre variabili di telemonitoraggio domestico è in corso.

I modelli ANOVA sono stati calcolati con il software STATISTICA, versione 10.0 (StatSoft Inc., www.statsoft.com). Nelle analisi statistiche ANOVA dei marker di rsEEG e MRI, il test di Mauchly sulla sfericità viene utilizzato per valutare se l'ipotesi di sfericità è soddisfatta, mentre la correzione di Greenhouse-Geisser viene applicata quando i dati violano tale ipotesi (Abdi, 2010). Il test di Duncan viene utilizzato per i confronti post-hoc (p < 0,05, correzione di Bonferroni).

Poiché l'uso di modelli ANOVA implica che la variabile dipendente deve essere normalmente distribuita, viene utilizzato il test di Kolmogorov-Smirnov (p < 0,05) per determinare se le caratteristiche di rsEEG, MRI e telemonitoraggio domestico di un dato modello ANOVA si avvicinano a distribuzioni gaussiane (ipotesi nulla di distribuzioni non gaussiane testata a p < 0,05). Questo prerequisito non è stato confermato in alcuni casi, quindi tutte le attività regionali della sorgente eLORETA sono state utilizzate come input per la trasformazione log 10 con il fine di rendere le soluzioni eLORETA gaussiane. Il test di Kolmogorov-Smirnov ha confermato che tutti i marker di rsEEG, MRI e del telemonitoraggio domestico sono gaussiani dopo questa trasformazione (p > 0,05).

La conferma dell'ipotesi di lavoro può richiedere (1) un'interazione ANOVA statisticamente significativa che includa il fattore Gruppo (p < 0,05) e (2) un test di Duncan post-hoc che indichi differenze statisticamente significative (p < 0,05, con correzione di Bonferroni) nei marker di rsEEG, MRI e di telemonitoraggio domestico tra i gruppi Nold e PDCD.

Analisi a livello individuale con gli strumenti di machine learning: accuratezza della classificazione dei partecipanti a Nold e con deficit cognitivo (PDCD ed MCI) in base ai marker selezionati di rsEEG, MRI e telemonitoraggio domestico.

Per l’analisi a livello individuale, utilizziamo strumenti di AI di machine learning per calcolare l'accuratezza della classificazioni dei partecipanti Nold e PDCD (MCI) sulla base dei marker di rsEEG, MRI e di telemonitoraggio domestico (caratteristiche) selezionati con l'analisi a livello di gruppo (le ANOVA di cui sopra) e arricchiti con informazioni demografiche sui partecipanti (età, sesso e anni di istruzione).

Per ogni strumento di machine learning implementato, usiamo un'analisi parallela dei dati e i risultati di tutti gli strumenti di machine learning utilizzati vengono mediati per produrre l'accuratezza media di classificazione tra partecipanti Nold e PDCD (MCI).

Per testare l'accuratezza della classificazione di ogni strumento di machine learning, i valori originali di ogni marker (e.g., di rsEEG, MRI e di telemonitoraggio domestico) selezionato sono proporzionalmente ri-scalati in un intervallo da 0 (valore minimo tra tutti i partecipanti Nold e pazienti con declino cognitivo lieve (MCI) dovuto a malattia neurodegenerativa (AD e PD) considerati nell’analisi) a 1 (valore massimo tra tutti i partecipanti Nold e ed MCI considerati nell’analisi). Ad esempio, per attività rsEEG nella banda di frequenza alpha 3, si procede come segue: (1) si considerano i valori di quel marker in tutti i partecipanti Nold e MCI dell’analisi come una singola popolazione; (2) il partecipante con il valore più basso di quel marker riceve il valore zero, il partecipante con il valore più alto riceve il valore 1, e gli altri partecipanti ricevono un valore proporzionale da 0 a 1. Per il marker MRI denominato volume normalizzato del DMN, si procede in modo analogo: (1) si considerano i valori di quel marker in tutti i partecipanti Nold e con deficit cognitivo (PDCD ed MCI) dell’analisi come una singola popolazione; (2) il partecipante con il valore più basso di quel marker riceve il valore zero, il partecipante con il valore più alto riceve il valore 1, e gli altri partecipanti ricevono un valore proporzionale da 0 a 1. Anche le variabili demografiche sono ri-scalate da 0 a 1. Ad esempio, per l’età dei partecipanti, si procede come segue: (1) si considerano i valori di quel marker in tutti i partecipanti Nold e con deficit cognitivo (PDCD ed MCI) dell’analisi come una singola popolazione; (2) il partecipante con il valore più basso di quel marker riceve il valore zero, il partecipante con il valore più alto riceve il valore 1, e gli altri partecipanti ricevono un valore proporzionale da 0 a 1. La variabile del genere sessuale è binaria e assume valore 0 nelle femmine e 1 nei maschi.

Riguardo il disegno di cross-validazione dei risultati del machine learning, utilizziamo la strategia della convalida incrociata a cinque volte con venti prove ciascuna per ridurre la distorsione introdotta dal rimescolamento casuale dei singoli set di dati utilizzati nel gruppo di addestramento (70% dei set di dati) e nel gruppo di test (30%). I risultati riportati in questo Deliverable 3.1. sono prodotti calcolando la media di quelli ottenuti dalle venti prove eseguite.

I seguenti indici percentuali rappresentano le prestazioni dei suddetti strumenti di machine learning nella predizione: (1) accuratezza come media tra sensibilità e specificità ponderata per il numero di partecipanti Nold e PDCD, (2) sensibilità per misurare il tasso di pazienti PDCD classificati correttamente; (3) specificità per misurare il tasso di partecipanti Nold classificati correttamente; (4) precisione per misurare il tasso di pazienti PDCD classificati correttamente sul totale (classificati e veri PDCD); e (5) F1-score, una variante dell’accuratezza che tiene conto della sensibilità e precisione. Si riportano i risultati relativi ai migliori 3 classificati per ogni strumento.

*Fase 2 dell'analisi statistica: associazione mediante machine learning tra i marker di funzione cognitiva globale (MMSE) e i marcatori (caratteristiche) di rsEEG, MRI e di telemonitoraggio domestico in partecipanti Nold e pazienti con declino cognitivo lieve (MCI) dovuto a malattia neurodegenerativa (AD e PD)*

Per verificare l’ipotesi che gli strumenti di machine learning possano prevedere con precisione i marker derivati dalla raccolta di dati effettuata in pazienti con declino cognitivo lieve (MCI) dovuto a malattia neurodegenerativa (AD e PD) in ambito ospedaliero sulla base dei marker demografici e di telemonitoraggio domestico, si usano i modelli di analisi di regressione di variabili continue di tali strumenti. Come per la discriminazione tra gruppi con deficit cognitivo (PDCD ed MCI) e Nold, utilizziamo la strategia della convalida incrociata a cinque volte con venti prove per ridurre la distorsione introdotta dal rimescolamento casuale dei singoli set di dati utilizzati nel gruppo di addestramento (70% dei set di dati) e nel gruppo di test (30%). Anche in questa fase dell’analisi statistica, i risultati riportati in questo D. 3.1. saranno calcolati facendo la media di quelli ottenuti dalle venti prove eseguite. Si riportano i risultati relativi ai migliori 3 classificati per ogni strumento, in termini di Mean Squared Error (ME), Root Mean Squared Error (RMSE) e Mean Absolute Error (MAE). Quest’ultimo fornisce una misura dell’accuratezza di predizione del modello, quantificando il tasso medio con cui il modello non predice correttamente i valori osservati. Per questa analisi di associazione, i partecipanti con deficit cognitivo (PDCD ed MCI) e Nold sono considerati come parte della stessa popolazione.

**Risultati**

*Analisi dei marker rsEEG*

Per i risultati dell’analisi quantitativa dei dati rsEEG, si veda il Deliverable 2.2. (“Estrazione dei marcatori EEG e relativa analisi statistica”). Ricapitolando, i marker rsEEG che hanno mostrato le differenze statisticamente significative tra i gruppi di partecipanti con PDCD e quelli Nold in base all’ANOVA (p < 0.0001) sono riportati in Tabella 4. Essi sono utilizzati come input rsEEG (features) per l’analisi con le tecniche di machine learning.

***Tabella 4***

|  |  |  |
| --- | --- | --- |
| delta F | theta P | alfa1 T |
| delta C | theta O | alfa2 O |
| delta P | theta T | alfa3 P |
| delta O | alfa1 F | alfa3 O |
| delta T | alfa1 C | delta globale |
| theta F | alfa1 P | theta globale |
| theta C | alfa1 O | alfa1 globale |
|  |  | alfa3 globale |

***Tabella 4.*** *Variabili rsEEG che hanno mostrato differenze statisticamente significative tra i gruppi di partecipanti con PDCD e quelli Nold in base all’ANOVA (p < 0.0001, si veda il Deliverable 2.2. per i dettagli).*

Sulla base dei suddetti marker rsEEG e quelli demografici, le tecniche di machine learning hanno mostrato i seguenti risultati di classificazione con accuratezza > 80% tra partecipanti con PDCD e quelli Nold (**Tabella 5**).

***Tabella 5***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classificatore** | **Accuratezza (%)** | **Sensibilità (%)** | **Specificità (%)** | **Precisione**  **(%)** | **F1\_score**  **(%)** |
| **DA 1** | 79.8 | 73.3 | 85.9 | 83.4 | 77.7 |
| **DA 2** | 79.8 | 71.7 | 87.5 | 85.3 | 77.2 |
| **DA3** | 79.5 | 80 | 79.1 | 78.9 | 78.8 |
| **GLRM 1** | 80.6 | 77.7 | 83.4 | 82.4 | 79.1 |
| **GLRM 2** | 80.5 | 81.7 | 79.4 | 79.1 | 80.2 |
| **GLRM 3** | 80.2 | 80.7 | 79.7 | 79.6 | 79.6 |
| **RF 1** | 82.4 | 81.3 | 83.4 | 82.6 | 81.5 |
| **RF 2** | 82.3 | 77.3 | 86.9 | 85.6 | 80.5 |
| **RF 3** | 82.3 | 79.3 | 85.0 | 84.2 | 81.0 |
| **K-NN 1** | 86.3 | 84.7 | 87.8 | 87.4 | 85.6 |
| **K-NN 2** | 85.2 | 83.3 | 86.9 | 85.9 | 84.3 |
| **K-NN 3** | 85.2 | 87.3 | 83.1 | 83.5 | 85.1 |
| **SVM 1** | 84.5 | 82.3 | 86.6 | 86.2 | 83.6 |
| **SVM 2** | 83.5 | 73.7 | 92.8 | 91.2 | 81.1 |
| **SVM 3** | 83.5 | 84.7 | 82.5 | 82.6 | 83.1 |

***Tabella 5.*** *Risultati dell’applicazione dei 5 diversi strumenti di* *machine learning al fine di classificare i soggetti con PDCD e quelli Nold sulla base dei marker rsEEG e demografici. Sono riportati i risultati relativi ai primi 3 classificatori migliori in termini di accuratezza nella classificazione per i 5 strumenti. Legenda: DA = Discriminant Analysis; GLRM = Generalized Linear Regression Model; RF = Random Forest; k-NN = K-nearest neighbor; SVM = Support Vector Machine.*

Analogamente. le tecniche di machine learning hanno mostrato i seguenti risultati di associazione con predizione > 80% tra i suddetti marker rsEEG e quelli demografici e il punteggio di MMSE non corretti nei partecipanti con PDCD e quelli Nold considerati come un'unica popolazione (**Tabella 6**).

***Tabella 6***

|  |  |  |  |
| --- | --- | --- | --- |
| **Classificatore** | **MSE** | **RMSE** | **MAE** |
| **GLRM 1** | 0.074257128 | 0.271419124 | 0.216745823 |
| **GLRM 2** | 0.075966848 | 0.274132291 | 0.218955274 |
| **GLRM 3** | 0.076556929 | 0.275686036 | 0.217130116 |
| **RF REG 1** | 0.057205306 | 0.238169018 | 0.190284917 |
| **RF REG 2** | 0.059162679 | 0.241414398 | 0.190985138 |
| **RF REG 3** | 0.06029789 | 0.243053753 | 0.187067114 |
| **SVM REG 1** | 0.062067053 | 0.246613106 | 0.185728315 |
| **SVM REG 2** | 0.061966125 | 0.247055716 | 0.189294936 |
| **SVM REG 3** | 0.062074973 | 0.247425532 | 0.191075969 |

***Tabella 6.*** *Risultati dell’applicazione dei 3 diversi strumenti di* *machine learning al fine di studiare l’associazione e la capacità di predizione del punteggio MMSE utilizzando in input i marker rsEEG e quelli demografici, considerando i soggetti con PDCD e quelli Nold come un unico gruppo. Sono riportati i risultati relativi ai primi 3 classificatori migliori in termini di accuratezza nella classificazione per i 3 strumenti. Legenda: GLRM = Generalized Linear Regression Model; RF = Random Forest; SVM = Support Vector Machine; MSE = Mean Squared Error; RMSE = Root Mean Squared Error; MAE = Mean Absolute Error.*

*Analisi dei marker MRI*

I marker MRI che hanno mostrato le differenze statisticamente significative tra i gruppi di partecipanti con MCI e quelli con PDCD in base all’ANOVA (p < 0.0001) sono riportati in **Tabella 7**. Essi sono utilizzati come input MRI (features) per l’analisi con le tecniche di machine learning. I marcatori di network non utilizzati nelle analisi riportate nel presente D. 3.1. (spessori nelle regioni dei network corticali, inclusi i gangli della base ed i marker globali di volume della materia grigia, bianca e di CSF), assai promettenti, richiedono una revisione dell’analisi dei dati i cui risultati saranno esposti nel Deliverable 5.1. “Rapporto su ottimizzazione algoritmi AI”.

***Tabella 7***

|  |  |
| --- | --- |
| Volume GM normalizzato  nel DMN | Volume GM normalizzato  nel FPN |
| Volume GM normalizzato  nel VN | Volume GM normalizzato  nel SMN |
| Volume GM normalizzato  nel DAN | Volume GM normalizzato  nel VAN |
| Volume GM normalizzato  nel LN |  |

***Tabella 7.*** *Variabili MRI che hanno mostrato differenze statisticamente significative tra i gruppi di partecipanti con PDCD e quelli Nold in base all’ANOVA (p < 0.0001). Legenda: GM = Grey Matter, volume di materia grigia; DMN = DEFAULT MODE NETWORK; FPN = FRONTOPARIETAL NETWORK; VISUAL NETWORK; SMN = SOMATOMOTOR NETWORK; DAN = DORSAL ATTENTION NETWORK; VAN = VENTRAL ATTENTION NETWORK; LN = LIMBIC NETWORK.*

Sulla base dei suddetti marker MRI e quelli demografici, le tecniche di machine learning hanno mostrato i seguenti risultati di classificazione con accuratezza > 80% tra partecipanti con PDCD e quelli Nold (**Tabella 8**).

***Tabella 8***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classificatore** | **Accuratezza (%)** | **Sensibilità (%)** | **Specificità (%)** | **Precisione**  **(%)** | **F1\_score**  **(%)** |
| **DA 1** | 72.2 | 73.1 | 71.0 | 77.0 | 74.2 |
| **DA 2** | 72.0 | 73.8 | 69.5 | 76.5 | 74.8 |
| **DA3** | 71.7 | 78.5 | 63.0 | 73.8 | 75.7 |
| **GLRM 1** | 74.3 | 80.0 | 67.0 | 76.5 | 77.9 |
| **GLRM 2** | 72.6 | 76.9 | 67.0 | 76.1 | 76.0 |
| **GLRM 3** | 72.4 | 70.8 | 74.5 | 79.9 | 74.2 |
| **RF 1** | 74.3 | 78.5 | 69.0 | 77.2 | 77.3 |
| **RF 2** | 74.3 | 78.5 | 69.0 | 77.5 | 77.1 |
| **RF 3** | 74.1 | 72.7 | 76.0 | 80.8 | 75.8 |
| **K-NN 1** | 78.7 | 81.2 | 75.5 | 81.8 | 80.9 |
| **K-NN 2** | 78.0 | 71.2 | 87.0 | 89.1 | 78.2 |
| **K-NN 3** | 77.6 | 77.3 | 78.0 | 83.6 | 79.7 |
| **SVM 1** | 78.5 | 78.8 | 78.0 | 83.7 | 80.4 |
| **SVM 2** | 78.0 | 80.8 | 74.5 | 80.8 | 80.3 |
| **SVM 3** | 77.8 | 79.2 | 76.0 | 82.4 | 80.1 |

***Tabella 8.*** *Risultati dell’applicazione dei 5 diversi strumenti di* *machine learning al fine di classificare i soggetti con PDCD e quelli Nold sulla base dei marker MRI e demografici. Sono riportati i risultati relativi ai primi 3 classificatori migliori in termini di accuratezza nella classificazione per i 5 strumenti. Legenda: DA = Discriminant Analysis; GLRM = Generalized Linear Regression Model; RF = Random Forest; k-NN = K-nearest neighbor; SVM = Support Vector Machine.*

Analogamente, le tecniche di machine learning hanno mostrato i seguenti risultati di associazione con accuratezza di predizione > 80% utilizzando i suddetti marker MRI e quelli demografici e il punteggio di MMSE non corretti nei partecipanti con PDCD e quelli Nold considerati come un'unica popolazione (**Tabella 9**). L’accuratezza di predizione è quantificata dal Mean Absolute Error (MAE), vale a dire il tasso medio con cui il modello non predice correttamente i valori osservati (vedi Metodi).

***Tabella 9***

|  |  |  |  |
| --- | --- | --- | --- |
| **Classificatore** | **MSE** | **RMSE** | **MAE** |
| **GLRM 1** | 0.05471373 | 0.231340837 | 0.180429481 |
| **GLRM 2** | 0.054925192 | 0.232905199 | 0.185392015 |
| **GLRM 3** | 0.056154679 | 0.235122822 | 0.186066443 |
| **RF REG 1** | 0.046469165 | 0.21236414 | 0.169927064 |
| **RF REG 2** | 0.047168759 | 0.214724067 | 0.168955176 |
| **RF REG 3** | 0.046796656 | 0.215119571 | 0.171692968 |
| **SVM REG 1** | 0.052102974 | 0.222211776 | 0.175120671 |
| **SVM REG 2** | 0.050821301 | 0.224250342 | 0.17596584 |
| **SVM REG 3** | 0.053566186 | 0.226396208 | 0.176280904 |

***Tabella 9.*** *Risultati dell’applicazione dei 3 diversi strumenti di* *machine learning al fine di studiare l’associazione e la capacità di predizione del punteggio MMSE utilizzando in input i marker MRI e quelli demografici, considerando i soggetti con PDCD e quelli Nold come un unico gruppo. Sono riportati i risultati relativi ai primi 3 classificatori migliori in termini di accuratezza nella predizione per i 3 strumenti. Legenda: GLRM = Generalized Linear Regression Model; RF = Random Forest; SVM = Support Vector Machine; MSE = Mean Squared Error; RMSE = Root Mean Squared Error; MAE = Mean Absolute Error.*

*Analisi dei marker di telemonitoraggio domestico delle funzioni cognitive (app SmartMe&You)*

Tutti i marker di telemonitoraggio domestico mostrano una differenze statisticamente significativa tra i gruppi di partecipanti con MCI e quelli Nold in base all’ANOVA, con un effetto statisticamente significativo del fattore Gruppo (Nold ≠ MCI, p < 0,001). Per tutti i task, i soggetti Nold hanno mostrato un’accuratezza superiore rispetto ai pazienti MCI. I valori di accuratezza di tutti i task sono utilizzati come input di telemonitoraggio (features) per l’analisi con le tecniche di machine learning.

Sulla base dei suddetti marker di telemonitoraggio domestico delle funzioni cognitive con l’app SmartMe&You e quelli demografici, le tecniche di machine learning hanno mostrato i seguenti risultati di classificazione con accuratezza > 80% tra partecipanti con MCI e quelli Nold (**Tabella 10**).

***Tabella 10***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classificatore** | **Accuratezza (%)** | **Sensibilità (%)** | **Specificità (%)** | **Precisione**  **(%)** | **F1\_score**  **(%)** |
| **DA 1** | 90.7 | 92.5 | 89.4 | 87.7 | 89.6 |
| **DA 2** | 90.4 | 87.5 | 92.5 | 90.5 | 88.3 |
| **DA3** | 90.0 | 90.0 | 90.0 | 88.5 | 88.4 |
| **GLRM 1** | 88.6 | 80.0 | 95.0 | 93.9 | 84.5 |
| **GLRM 2** | 88.2 | 82.5 | 92.5 | 90.3 | 85.4 |
| **GLRM 3** | 88.2 | 78.3 | 95.6 | 93.6 | 84.2 |
| **RF 1** | 92.9 | 91.4 | 94.3 | 94.5 | 92.5 |
| **RF 2** | 92.9 | 91.4 | 94.3 | 95.0 | 92.6 |
| **RF 3** | 92.5 | 92.9 | 92.1 | 92.8 | 92.4 |
| **K-NN 1** | 95.0 | 94.3 | 95.7 | 95.9 | 94.8 |
| **K-NN 2** | 94.6 | 95.0 | 94.3 | 95.0 | 94.7 |
| **K-NN 3** | 94.6 | 93.6 | 95.7 | 96.2 | 94.3 |
| **SVM 1** | 94.3 | 94.3 | 94.3 | 95.0 | 94.2 |
| **SVM 2** | 93.9 | 95.0 | 92.9 | 93.5 | 94.0 |
| **SVM 3** | 93.6 | 94.3 | 92.9 | 93.6 | 93.6 |

***Tabella 10.*** *Risultati dell’applicazione dei 5 diversi strumenti di* *machine learning al fine di classificare i soggetti con MCI e Nold sulla base dei marker di telemonitoraggio domestico e demografici. Sono riportati i risultati relativi ai primi 3 classificatori migliori in termini di accuratezza nella classificazione per i 5 strumenti. Legenda: DA = Discriminant Analysis; GLRM = Generalized Linear Regression Model; RF = Random Forest; k-NN = K-nearest neighbor; SVM = Support Vector Machine.*

Analogamente, le tecniche di machine learning hanno mostrato i seguenti risultati di associazione con accuratezza di predizione> 80% utilizzando i suddetti marker di telemonitoraggio domestico delle funzioni cognitive con l’app SmartMe&You e quelli demografici rispetto al punteggio di MMSE non corretti nei partecipanti con MCI e quelli Nold considerati come un'unica popolazione (**Tabella 11**).

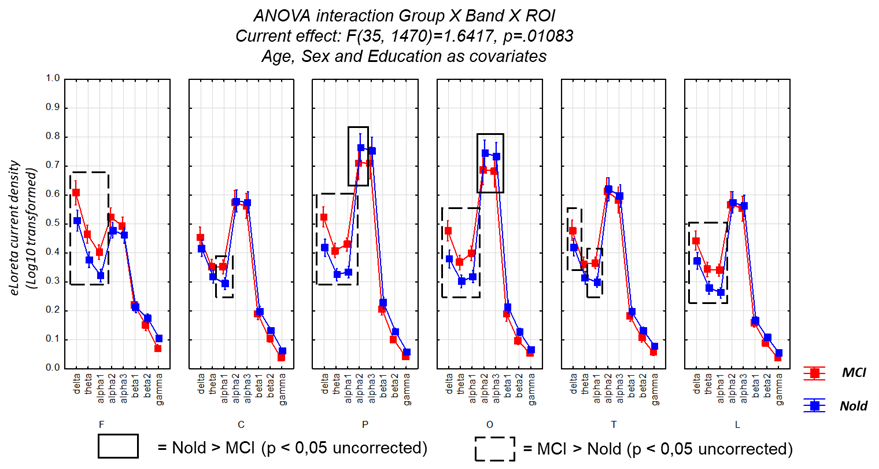
***Tabella 11***

|  |  |  |  |
| --- | --- | --- | --- |
| **Regressore** | **MSE** | **RMSE** | **MAE** |
| **GLRM 1** | 0.058701083 | 0.23738409 | 0.176330498 |
| **GLRM 2** | 0.063261377 | 0.243262223 | 0.184379024 |
| **GLRM 3** | 0.0649144707 | 0.246771677 | 0.1892668014 |
| **RF REG 1** | 0.067406744 | 0.250464653 | 0.187226828 |
| **RF REG 2** | 0.066270848 | 0.251114755 | 0.186988636 |
| **RF REG 3** | 0.067903442 | 0.252924968 | 0.171915278 |
| **SVM REG 1** | 0.067406744 | 0.250464653 | 0.187226828 |
| **SVM REG 2** | 0.066270848 | 0.251114755 | 0.186988636 |
| **SVM REG 3** | 0.067903442 | 0.252924968 | 0.171915278 |

***Tabella 11.*** *Risultati dell’applicazione dei 3 diversi strumenti di* *machine learning al fine di studiare l’associazione e la capacità di predizione del punteggio MMSE utilizzando in input i marker di telemonitoraggio e quelli demografici, considerando i soggetti con MCI e Nold come un unico gruppo. Sono riportati i risultati relativi ai primi 3 classificatori migliori in termini di accuratezza nella predizione per i 3 strumenti. Legenda: GLRM = Generalized Linear Regression Model; RF = Random Forest; SVM = Support Vector Machine; MSE = Mean Squared Error; RMSE = Root Mean Squared Error; MAE = Mean Absolute Error.*

In linea con risultati precedenti, l’analisi dei dati rsEEG relativi al presente gruppo di partecipanti con MCI e quelli Nold mostra un'interazione ANOVA statisticamente significativa (p < 0,01) tra i fattori Gruppo (Nold, MCI), Banda (delta, theta, alfa1, alfa2, alfa3, beta1, beta2, gamma), e Regione (frontale, centrale, parietale, occipitale, temporale, e limbica) . La Figura 1 illustra questi risultati.

***Figura 1***



***Figura 1.*** *Valori medi (± errore standard della media, SE, trasformati Log10) delle attività di sorgente rsEEG (densità di corrente eLORETA normalizzata) relativi a un effetto di interazione ANOVA statisticamente significativo (F(35, 1470) = 1,6417, p = 0,01083) tra i fattori Gruppo (Nold, N = 27; e MCI, N = 20), Regioni di interesse (frontale, centrale, parietale, occipitale, temporale e limbico) e Banda (delta, theta, alfa 1, alfa 2, alfa 3, beta 1 e beta 2). Come covariate sono stati utilizzate l’età, l’istruzione ed il sesso. I rettangoli indicano le bande di frequenza in cui le soluzioni eLORETA presentano una differenza statisticamente significativa tra i gruppi Nold e MCI (p < 0,05 non corretto). Legenda: Nold = anziano cognitivamente integro; MCI = deficit cognitivo lieve; rsEEG = ritmi elettroencefalografici in condizione di veglia rilassata ad occhi chiusi.*

Le tecniche di machine learning hanno mostrato i seguenti risultati di associazione con accuratezza nella predizione > 80% utilizzando i suddetti marker di telemonitoraggio domestico delle funzioni cognitive con l’app SmartMe&You e quelli demografici rispetto al marker della sorgente rsEEG theta parietale, risultata più rappresentativa della differenza tra i gruppi, nei partecipanti con MCI e quelli Nold considerati come un'unica popolazione (**Tabella 13**).

***Tabella 12***

|  |  |  |  |
| --- | --- | --- | --- |
| **Regressore** | **MSE** | **RMSE** | **MAE** |
| **GLRM 1** | 0.086892778 | 0.291719237 | 0.237917773 |
| **GLRM 2** | 0.090950672 | 0.298994595 | 0.243712938 |
| **GLRM 3** | 0.090760088 | 0.299154323 | 0.244984799 |
| **RF 1** | 0.059132295 | 0.241486126 | 0.185896989 |
| **RF 2** | 0.060526161 | 0.244733713 | 0.193962211 |
| **RF 3** | 0.0605281 | 0.244906951 | 0.193427671 |
| **SVM 1** | 0.072552443 | 0.267427747 | 0.215986135 |
| **SVM 2** | 0.072521005 | 0.267731905 | 0.209774875 |
| **SVM 3** | 0.072776023 | 0.268696969 | 0.214273006 |

***Tabella 12.*** *Risultati dell’applicazione dei 3 diversi strumenti di* *machine learning al fine di studiare l’associazione e la capacità di predizione della variabile rsEEG theta parietale utilizzando in input i marker di telemonitoraggio e quelli demografici, considerando i soggetti con MCI e Nold come un unico gruppo. Sono riportati i risultati relativi ai primi 3 classificatori migliori in termini di accuratezza nella predizione per i 3 strumenti. Legenda: GLRM = Generalized Linear Regression Model; RF = Random Forest; SVM = Support Vector Machine; MSE = Mean Squared Error; RMSE = Root Mean Squared Error; MAE = Mean Absolute Error.*

**Conclusioni**

In questo Deliverable 3.1. del Progetto TELEMAIA, abbiamo descritto i risultati dell’uso delle procedure di machine learning usate per la classificazione tra partecipanti Nold e quelli con deficit cognitivo (PDCD ed MCI) sulla base dei marker di rsEEG, MRI e di telemonitoraggio domestico delle funzioni cognitive con l’app “SmartMe&You” e lo studio di associazione tra i suddetti marker di telemonitoraggio domestico delle funzioni cognitive (arricchiti con i marker demografici) e i marker di rsEEG delle funzione cerebrale e i punteggi di mini-mental state examination (MMSE) che riflettono lo stato cognitivo globale.

I risultati del modello ANOVA hanno mostrato differenze per sorgenti rsEEG (vedi Deliverable 2.2), MRI e del telemonitoraggio domestico delle funzioni cognitive con l’app SmartMe&You tra gruppi MCI e Nold dell’ordine di p < 0.000001. La capacità di discriminazione tra i partecipanti con deficit cognitivo (PDCD ed MCI) e quelli Nold di tali marker, valutata con cinque diverse tecniche di machine learning, ha dato valori di accuratezza di classificazione > dell’ 80%.

Inoltre, i marker di telemonitoraggio domestico (arricchiti dai marker demografici) hanno mostrato una capacità di predizione del punteggio di MMSE, valutata con tre diverse tecniche di machine learning, nei partecipanti con deficit cognitivo (PDCD ed MCI) e quelli Nold, considerati come un’unica popolazione, con errore nella predizione quantificato dal Mean Absolute Error (MAE) < del 20%. Tali risultati sono stati anche ottenuti anche per la predizione del punteggio di MMSE utilizzando i marker rsEEG e MRI (arricchiti dai marker demografici) in analisi separate, in linea con risultati precedenti di letteratura. Infine, tali accuratezze di predizione medio-alte sono state raggiunte anche utilizzando i marker di telemonitoraggio domestico (arricchiti dai marker demografici) per predire la variabile rsEEG (theta parietale). Da sottolineare è la coerenza dei valori di accuratezza ottenuti sia per la classificazione che per l’associazione (predizione) tra le varie tecniche di machine learning, evidenziando la robustezza dei risultati.

La procedura sperimentale per la selezione e l’uso dei marker di rsEEG, MRI e di telemonitoraggio domestico delle funzioni cognitive con l’app SmartMe&You presentata in questo Deliverable 3.1. ha pienamente mostrato la sua efficacia per la discriminazione statistica di pazienti con declino cognitivo rispetto a partecipanti Nold. Essa ha anche mostrato che le prestazioni del telemonitoraggio domestico delle funzioni cognitive con l’app SmartMe&You sono associate allo stato cognitivo globale rivelato dai punteggio di MMSE e alle disfunzioni cerebrali riflesse nelle variabili di rsEEG nella popolazione di pazienti e controlli di questo studio. I risultati ottenuti potrebbero essere applicati su altre patologie neurodegenerative, quali la malattia di Alzheimer (AD) promuovendo l’uso del telemonitoraggio.

Le tecniche di machine learning sviluppate nel D. 3.1 e testate con successo sui marker tesi marker EEG, MRI e di telemonitoraggio domestico di funzioni cognitivo-motorie derivati da pazienti con declino cognitivo dovuto a malattie neurodegenerative (malattie di Parkinson e Alzheimer) e gli stessi marker saranno oggetto di un lavoro intenso di ottimizzazione nelle attività progettuali del WP5 del Progetto TELEMAIA e i risultati di tali attività saranno esposti nel D.5.1.

**Bibliografia**

Adeli E, Shi F, An L, Wee CY, Wu G, Wang T, Shen D. Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data. Neuroimage. 2016 Nov 1;141:206-219. doi: 10.1016/j.neuroimage.2016.05.054. Epub 2016 Jun 10. PMID: 27296013.

Alloni A, Quaglini S, Panzarasa S, Sinforiani E, Bernini S. Evaluation of an ontology-based system for computerized cognitive rehabilitation. Int J Med Inform. 2018;115:64-72. doi: 10.1016/j.ijmedinf.2018.04.005. PMID: 29779721.

Babiloni C, Del Percio C, Lizio R, Noce G, Cordone S, Lopez S, Soricelli A, Ferri R, Pascarelli MT, Nobili F, Arnaldi D, Famà F, Aarsland D, Orzi F, Buttinelli C, Giubilei F, Onofrj M, Stocchi F, Stirpe P, Fuhr P, Gschwandtner U, Ransmayr G, Caravias G, Garn H, Sorpresi F, Pievani M, D&#39;Antonio F, De Lena C, Güntekin B, Hanoğlu L, Başar E, Yener G, Emek-Savaş DD, Triggiani AI, Franciotti R, Frisoni GB, Bonanni L, De Pandis MF. Abnormalities of Cortical Neural Synchronization Mechanisms in Subjects with Mild Cognitive Impairment due to Alzheimer&#39;s and Parkinson&#39;s Diseases: An EEG Study. J Alzheimers Dis. 2017;59(1):339-358. doi: 10.3233/JAD-160883. PMID: 28621693.

Babiloni C, Del Percio C, Lizio R, Noce G, Lopez S, Soricelli A, Ferri R, Pascarelli MT, Catania V, Nobili F, Arnaldi D, Famà F, Orzi F, Buttinelli C, Giubilei F, Bonanni L, Franciotti R, Onofrj M, Stirpe P, Fuhr P, Gschwandtner U, Ransmayr G, Garn H, Fraioli L, Pievani M, D&#39;Antonio F, De Lena C, Güntekin B, Hanoğlu L, Başar E, Yener G, Emek-Savaş DD, Triggiani AI, Taylor JP, De Pandis MF, Vacca L, Frisoni GB, Stocchi F. Functional cortical source connectivity of resting state electroencephalographic alpha rhythms shows similar abnormalities in patients with mild cognitive impairment due to Alzheimer’s and Parkinson’s diseases. Clin Neurophysiol. 2018a;129(4):766-782. doi: 10.1016/j.clinph.2018.01.009. PMID: 29448151.

Babiloni C, Del Percio C, Lizio R, Noce G, Lopez S, Soricelli A, Ferri R, Pascarelli MT, Catania V, Nobili F, Arnaldi D, Famà F, Aarsland D, Orzi F, Buttinelli C, Giubilei F, Onofrj M, Stocchi F, Vacca L, Stirpe P, Fuhr P, Gschwandtner U, Ransmayr G, Garn H, Fraioli L, Pievani M, Frisoni GB, D&#39;Antonio F, De Lena C, Güntekin B, Hanoğlu L, Başar E, Yener G, Emek-Savaş DD, Triggiani AI, Franciotti R, Taylor JP, De Pandis MF, Bonanni L. Abnormalities of Resting State Cortical EEG Rhythms in Subjects with Mild Cognitive Impairment Due to Alzheimer&#39;s and Lewy Body Diseases. J Alzheimers Dis. 2018b;62(1):247-268. doi: 10.3233/JAD-170703. PMID: 29439335.

Babiloni C, Del Percio C, Lizio R, Noce G, Lopez S, Soricelli A, Ferri R, Pascarelli MT, Catania V, Nobili F, Arnaldi D, Famà F, Orzi F, Buttinelli C, Giubilei F, Bonanni L, Franciotti R, Onofrj M, Stirpe P, Fuhr P, Gschwandtner U, Ransmayr G, Fraioli L, Parnetti L, Farotti L, Pievani M, D&#39;Antonio F, De Lena C, Güntekin B, Hanoğlu L, Yener G, Emek-Savaş DD, Triggiani AI, Taylor JP, McKeith I, Stocchi F, Vacca L, Frisoni GB, De Pandis MF. Levodopa may affect cortical excitability in Parkinson&#39;s disease patients with cognitive deficits as revealed by reduced activity of cortical sources of resting state electroencephalographic rhythms. Neurobiol Aging. 2019;73:9-20. doi: 10.1016/j.neurobiolaging.2018.08.010. PMID: 30312790.

Babiloni C, Pascarelli MT, Lizio R, Noce G, Lopez S, Rizzo M, Ferri R, Soricelli A, Nobili F, Arnaldi D, Famà F, Orzi F, Buttinelli C, Giubilei F, Salvetti M, Cipollini V, Bonanni L, Franciotti R, Onofrj M, Stirpe P, Fuhr P, Gschwandtner U, Ransmayr G, Aarsland D, Parnetti L, Farotti L, Marizzoni M, D&#39;Antonio F, De Lena C, Güntekin B, Hanoğlu L, Yener G, Emek-Savaş DD, Triggiani AI, Taylor JP, McKeith I, Stocchi F, Vacca L, Hampel H, Frisoni GB, De Pandis MF, Del Percio C. Abnormal cortical neural synchronization mechanisms in quiet wakefulness are related to motor deficits, cognitive symptoms, and visual hallucinations in Parkinson&#39;s disease patients: an electroencephalographic study. Neurobiol Aging. 2020;91:88-111. doi: 10.1016/j.neurobiolaging.2020.02.029. PMID: 32234263.

Dauvergne C, Bégel V, Gény C, Puyjarinet F, Laffont I, Dalla Bella S. Home-based training of rhythmic skills with a serious game in Parkinson&#39;s disease: Usability and acceptability. Ann Phys Rehabil Med. 2018;61(6):380-385. doi: 10.1016/j.rehab.2018.08.002. PMID: 30193992.

De Luca R, Latella D, Maggio MG, Di Lorenzo G, Maresca G, Sciarrone F, Militi D, Bramanti P, Calabrò RS. Computer assisted cognitive rehabilitation improves visuospatial and executive functions in Parkinson&#39;s disease: Preliminary results. NeuroRehabilitation. 2019;45(2):285-290. doi: 10.3233/NRE-192789. PMID: 31498141.

de Oliveira PCA, de Araújo TAB, Machado DGDS, Rodrigues AC, Bikson M, Andrade SM, Okano AH, Simplicio H, Pegado R, Morya E. Transcranial Direct Current Stimulation on Parkinson&#39;s Disease: Systematic Review and Meta-Analysis. Front Neurol. 2022 Jan 10;12:794784. doi: 10.3389/fneur.2021.794784. PMID: 35082749; PMCID: PMC8785799.

Eusebi P, Franchini D, De Giorgi M, Abraha I, Montedori A, Casucci P, Calabresi P, Tambasco N. Incidence and prevalence of Parkinson&#39;s disease in the Italian region of Umbria: a population-based study using healthcare administrative databases. Neurol Sci. 2019;40(8):1709-1712. doi: 10.1007/s10072-019-03872-w. PMID: 30945089.

Gelb DJ, Oliver E, Gilman S. Diagnostic criteria for Parkinson disease. Arch Neurol. 1999;56(1):33-9. doi: 10.1001/archneur.56.1.33. PMID: 9923759.

Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson&#39;s disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry. 1992;55(3):181-4. doi: 10.1136/jnnp.55.3.181. PMID: 1564476; PMCID: PMC1014720.

Maggio MG, De Luca R, Maresca G, Di Lorenzo G, Latella D, Calabro RS, Bramanti A. Personal computer-based cognitive training in Parkinson&#39;s disease: a case study. Psychogeriatrics. 2018;18(5):427-429. doi: 10.1111/psyg.12333. PMID: 29987868.

Malerba F, Arisi I, Florio R, Zecca C, Dell&#39;Abate MT, Bruni Ercole B, Camerini S, Casella M, Logroscino G, Cattaneo A. proNGF Measurement in Cerebrospinal Fluid Samples of a Large Cohort of Living Patients With Alzheimer&#39;s Disease by a New Automated Immunoassay. Front Aging Neurosci. 2021;13:741414. doi: 10.3389/fnagi.2021.741414. PMID: 34776928; PMCID: PMC8579109.

Pang H, Wang J, Yu Z, Yu H, Li X, Bu S, Zhao M, Jiang Y, Liu Y, Fan G. Glymphatic function from diffusion-tensor MRI to predict conversion from mild cognitive impairment to dementia in Parkinson&#39;s disease. J Neurol. 2024. doi: 10.1007/s00415-024-12525-8. PMID: 38913186.

Teipel S, Gustafson D, Ossenkoppele R, Hansson O, Babiloni C, Wagner M, Riedel-Heller SG, Kilimann I, Tang Y. Alzheimer Disease: Standard of Diagnosis, Treatment, Care, and Prevention. J Nucl Med. 2022;63(7):981-985. doi: 10.2967/jnumed.121.262239. PMID: 35145015; PMCID: PMC9258577.

Zimmermann R, Gschwandtner U, Benz N, Hatz F, Schindler C, Taub E, Fuhr P. Cognitive training in Parkinson disease: cognition-specific vs nonspecific computer training. Neurology. 2014;82(14):1219-26. doi: 10.1212/WNL.0000000000000287.

**APPENDICE SULLE TECNICHE DI MACHINE LEARNING**

Il morbo di Parkinson è un disturbo neurologico complesso e debilitante che colpisce milioni di persone in tutto il mondo. È caratterizzata da una serie di sintomi motori e non motori, tra cui tremori, rigidità, bradicinesia, instabilità posturale e deterioramento cognitivo. La diagnosi precoce e accurata della malattia di Parkinson è fondamentale per un intervento medico tempestivo e per migliorare la qualità della vita del paziente. Attualmente, non esiste un test diagnostico definitivo per la malattia di Parkinson e la diagnosi viene in genere effettuata attraverso una combinazione di valutazione clinica, anamnesi e test di laboratorio. L'elettroencefalografia (EEG) è una tecnica di neuroimaging non invasiva che misura l'attività elettrica del cervello; in particolare è una tecnica di registrazione delle attività elettriche del cervello mediante l'uso di elettrodi posizionati sul cuoio capelluto e misura le variazioni di potenziale elettrico derivanti dall'attività neuronale e le rappresenta sotto forma di onde cerebrali. In alcune terapie, come la stimolazione magnetica transcranica (TMS,” Transcranial Magnetic Stimulation”) o la stimolazione cerebrale profonda (DBS, “Deep Brain Stimulation”), l'EEG può essere utilizzato per monitorare l'attività cerebrale e adattare i parametri di stimolazione per ottimizzare i benefici terapeutici. L'EEG può essere integrato con tecnologie di biofeedback per aiutare i pazienti a migliorare il controllo delle loro funzioni cerebrali attraverso esercizi di riabilitazione cognitiva. Di conseguenza, l'EEG è uno strumento fondamentale nella diagnosi, monitoraggio e ricerca delle malattie neurodegenerative, offrendo un'analisi dettagliata delle attività cerebrali che può guidare il trattamento e migliorare la comprensione di queste complesse condizioni.

È stato dimostrato che i segnali EEG sono alterati negli individui con malattia di Parkinson e possono fornire un prezioso biomarcatore per la diagnosi e il monitoraggio della malattia.

La classificazione in ambito scientifico e non solo viene utilizzata per assegnare elementi a un gruppo o a una classe in base a un set specifico di caratteristiche.

Il Machine Learning, o ML, è un sottocampo dell'Intelligenza Artificiale che ruota attorno allo sviluppo di algoritmi informatici sfruttando i dati i quali facilitano le macchine a prendere decisioni o previsioni analizzando e facendo inferenze dai dati. Come gli esseri umani acquisiscono conoscenza comprendendo gli input, l'apprendimento automatico mira a prendere decisioni a partire dai dati di input. La potenza dell'apprendimento automatico sono gli algoritmi. È possibile utilizzare diversi algoritmi di apprendimento automatico a seconda della struttura e del volume dei dati. Gli algoritmi ML sono un tipo di intelligenza artificiale che può essere utilizzata per analizzare i segnali EEG e identificare modelli o caratteristiche associati al morbo di Parkinson. L'EEG può catturare i biomarcatori elettrofisiologici associati al Parkinson, rilevare i sintomi non motori e offrire una diagnosi precoce. L'integrazione delle tecniche di apprendimento automatico con i dati EEG si è dimostrata promettente nell'identificazione di modelli e caratteristiche complessi associati al Parkinson, migliorando l'accuratezza diagnostica. La intelligenza artificiale (AI) mostra grandi potenzialità nella diagnosi del morbo di Parkinson (PD) anche a partire dalle scansioni MRI. Le scansioni MRI possono fornire preziose informazioni sulle differenze cerebrali strutturali e funzionali tra gli individui con PD e quelli senza; in tale contesto gli algoritmi di intelligenza artificiale possono essere addestrati per rilevare modelli sottili e anomalie indicativi di PD con elevata precisione ed efficienza: le applicazioni dell'AI in neurologia, tra cui la diagnosi, la prognosi e il trattamento delle malattie, nonché la medicina di precisione e la telemedicina devono porsi come obiettivo lo sviluppo di un modello di intelligenza artificiale in grado di distinguere tra le scansioni MRI di individui con PD e quelli senza, fornendo ai neurologi un robusto strumento diagnostico.

Il Deep Learning o DL è un sottocampo del Machine Learning che sfrutta le reti neurali per replicare il funzionamento di un cervello umano sulle macchine: i neuroni sono configurati nelle reti neurali in base all'addestramento su grandi quantità di dati (proprio come gli algoritmi sono i motori del Machine Learning, il Deep Learning ha dei modelli). Un modello di Deep Learning è costituito da nodi che formano diversi strati nella rete neurale: le informazioni vengono passate attraverso ogni strato della rete, la quale cerca di comprendere le informazioni e identificare i pattern nei dati.

Il DL garantisce un alto livello di precisione, ma richiede grandi quantità di dati per effettuare previsioni accurate, oltre che dell’hardware potente da sfruttare in fase di training in modo da velocizzare i tempi.

Esistono due approcci per utilizzare il DL per la classificazione delle immagini, il primo prevede la creazione da zero della rete (progettazione dell’architettura, scelta del dataset con immagini etichettate ecc.), mentre il secondo prevede l’utilizzo di modelli pre-addestrati (approccio “Transfer Learning”), ossia parte da reti pre-esistenti come ad esempio AlexNet o GoogleNet (tale approccio risulta più rapido del precedente) da addestrare sui nuovi dataset per il nuovo problema sotto esame.

I modelli di Deep Learning, come le reti neurali convoluzionali (CNN, Convolutional Neural Network), vengono utilizzati per imparare automaticamente le caratteristiche di un oggetto al fine di indentificarlo a differenza dell’approccio ML in cui occorre selezionare le features (vettori caratteristici che rappresentano le immagini in input) da utilizzare all’interno del sistema durante la fase di training come mostrato in Figura 1. In particolare le CNN sono utilizzate specialmente nel settore dell’elaborazione delle immagini.

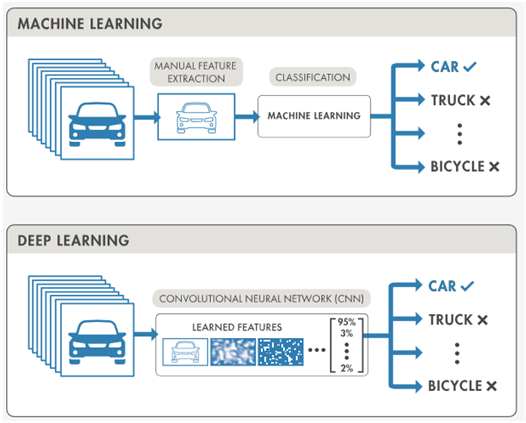


Figura 1: Differenza tra Machine Learning e Deep Learning

La classificazione nel nostro caso è un tipo di apprendimento automatico supervisionato (“supervised”) in cui un algoritmo "impara" a classificare nuove osservazioni da esempi di dati etichettati.

In pratica il modello addestrato tenta di stimare l'etichetta (label) corretta di un determinato dato di input. Nella classificazione, il modello viene sottoposto a training completo usando i dati di training e quindi viene valutato sui dati di test prima di essere usato per eseguire la stima su nuovi dati non ancora visualizzati. Un classico esempio è un algoritmo che può imparare a prevedere se una determinata e-mail è spam o meno (in questo caso si parla di classificazione binaria in quanto sono previste due possibili classi).

L'apprendimento non supervisionato (“unsupervised”) invece è un tipo di algoritmo di apprendimento automatico utilizzato per trarre inferenze da set di dati senza l'intervento umano; a differenza dell'apprendimento supervisionato in cui le etichette (label) vengono fornite insieme ai dati. Il metodo di apprendimento non supervisionato più comune è l'analisi dei cluster, che applica metodi di clustering per esplorare i dati e trovare pattern o raggruppamenti nascosti nei dati (l’algoritmo più popolare in tale ambito è il k-Means).

Ricapitolando ’utilizzo del Machine Learning per il riconoscimento di oggetti offre la flessibilità di scegliere la migliore combinazione di features e classificatori per l’apprendimento consentendo di otte-nere risultati precisi con una piccola quantità di dati. Gli algoritmi di apprendimento di ML come detto si dividono in due rami:

* “supervised” ossia supervisionati: il modello viene addestrato con dati in input/output noti in modo da imparare a riconoscere dati futuri (esempio: classificazione, regressione).
* “unsupervised” ossia non supervisionati: il modello riconosce automaticamente strutture intrinseche nascoste all’interno dei dati in ingresso (esempio: clustering).

Le metodiche feature-based estraggono le features (o vettore caratteristico) a partire dai target di interesse che dipendono dal problema sotto esame, ad esempio da immagini o segnali di varia natura (dopo eventuale pre-processing); tali feature vengono utilizzate per addestrare un opportuno classificatore in modalità offline.

Un esempio di estrazione di un vettore caratteristico a partire da una immagine (in questo caso SAR: “Synthetic Aperture Radar”) è riportato di seguito in Figura 2:

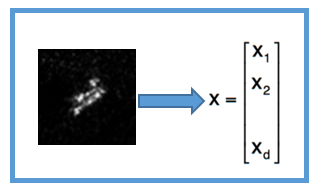


Figura 2: Estrazione Vettore Caratteristico

Le performance dipendono dalla scelta delle features in fase di addestramento e la capacità di queste di rappresentare in maniera efficiente i target.

Sono presenti in Letteratura diversi algoritmi di classificazione all’interno del mondo della Intelligenza Artificiale (I.A.) oltre che numerosi metodi per estrarre le features.

Nella fase di classificazione online, il classificatore addestrato prende in input le features estratte dai target ed assegna una label di appartenenza ad una delle classi con cui è stato addestrato precedentemente.

La decisione presa dal classificatore si basa puramente sulla somiglianza tra i vettori caratteristici estratti e quelli usati in fase di training.

Tale approccio, molto impiegato nella pratica come è possibile constatare dalla mole di pubblicazioni presenti in Letteratura, può andare incontro a dei problemi quando si presentano delle variazioni sostanziali nelle features estratte.

Un sistema di tipo Feature-based può essere visto come un insieme di funzioni di discriminazione date da:

in cui

* **x** rappresenta il vettore caratteristico di lunghezza D
* C rappresenta il numero di classi target.

Il classificatore assegna il vettore caratteristico in input alla classe qualora si ha:

di conseguenza il classificatore calcola C funzioni discriminanti e seleziona la classe con discriminate più alto (Figura 3).

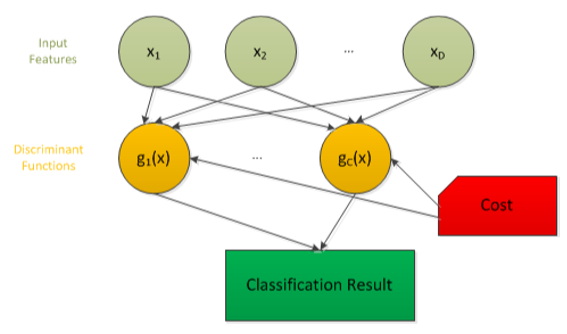


Figura 3: Schema Classificatore

L’operazione descritta equivale a partizionare lo spazio delle feature in C regioni di decisione:

in modo che quando si ha:

allora **x** appartiene alla regione e **x** è assegnato alla ipotesi .

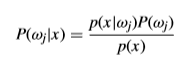
Di seguito saranno descritti i principali classificatori impiegati nelle applicazioni reali.

Per “ensemble learning” si intende l’utilizzo combinato di più classificatori addestrati individualmente in modo da migliorare le prestazioni rispetto al singolo classificatore soprattutto se ogni classificatore è altamente addestrato (“expert”) in una regione differente dello spazio delle feature. Ci sono diversi modi per comporre l’insieme tra cui “sub-sumpling-based” “mixture of experts” e “stacking”. Nel metodo sub-sampling-based i classificatori individuali sono addestrati su sotto-campioni di un dataset di training comune contenente le feature. Il ricampionamento avviene mediante algoritmi di “bagging” o “boosting” (ad esempio AdaBoost). Nel metodo “mixture of experts” il dataset di training è partizionato in diverse regioni ognuna delle quali viene assegnata ad un classificatore dell’insieme. Nel metodo “stacking” si impiega un “expert” di secondo livello per accettare l’output da un insieme di expert.

**BAYES CLASSIFICATION**

Il classificatore di Bayes si basa sulla teoria della decisione Bayesiana.

Date le C classi di target ognuna con funzione di densità di probabilità (PDF) e probabilità a priori , si ha che la probabilità a posteriori è data da:



in cui la PDF di x è data da:



L’espressione di prevede che la PDF abbia una forma nota e i vettori caratteristici vengono utilizzati per stimare i parametri della distribuzione; ad esempio nel caso di distribuzione Gaussiana la stima dei parametri della distribuzione (media e matrice di covarianza) per ogni classe a partire dai feature vector porta ad un classificatore quadratico ossia la cui superficie di demarcazione è quadratica (Figura 4). Tale classificatore presenta una matrice di covarianza S differente per ogni classe ().

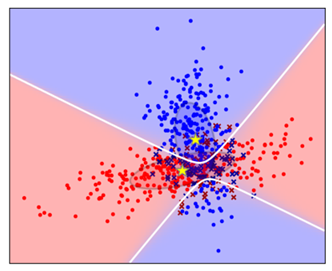


Figura 4: Classificatore Quadratico

Il classificatore Bayesiano può essere semplificato sotto alcune ipotesi. Nel caso in cui tutte le classi hanno eguale probabilità e presentano una matrice di covarianza S diagonale con tutti gli elementi lungo la diagonale uguali ossia:

in cui **I** rappresenta la matrice identità mentre σ2 la varianza, si parla di Classificatore a Minima Distanza Euclidea. Quando le probabilità a priori sono uguali e la matrice di covarianza risulta essere non diagonale si parla di Classificatore a Minima Distanza di Mahalanobis.

Nel caso in cui la matrice di covarianza risulta diagonale si parla di Classificatore Lineare.

Nel caso in cui le classi hanno una differente matrice di covarianza proporzionale alla matrice identità si ha un classificatore quadratico.

Naive Bayes è un algoritmo probabilistico di apprendimento automatico che può essere utilizzato in un'ampia varietà di attività di classificazione. Il nome naïve viene utilizzato perché presuppone che le feature inserite nel modello siano indipendenti l'una dall'altra. Ovvero la modifica del valore di una caratteristica, non influenza o modifica direttamente il valore di nessuna delle altre caratteristiche utilizzate nell'algoritmo.

Il Naive Bayes è una specifica implementazione del classificatore bayesiano che assume che le features del vettore caratteristico siano condizionalmente indipendenti tra loro data la classe. Questo è un forte assunto di indipendenza che spesso non è vero nella pratica, ma semplifica notevolmente i calcoli.

Un feature vector **x** può essere rappresentato come:

dove:

​sono le caratteristiche (features) o variabili predittive.

Nel Naive Bayes, si assume che ogni caratteristica ​ sia indipendente dalle altre caratteristiche ​ (per ) dato che conosciamo la classe ​. Questa è l'assunzione di indipendenza condizionata.

Questa assunzione semplifica notevolmente il calcolo della probabilità condizionata di **x** dato , poiché è possibile trattare ogni separatamente.

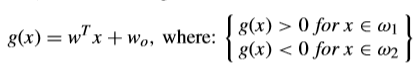
**DISCRIMINANT ANALYSIS**

I metodi di classificazione basati sulla Discriminant Analysis (DA) utilizzano funzioni discriminanti (DF, Discriminant Function); la DA è una tecnica statistica utilizzata per classificare un insieme di osservazioni in categorie o gruppi predefiniti; in particolare essa è un metodo di classificazione che presuppone che classi diverse generino dati basati su distribuzioni gaussiane diverse.

* Per addestrare (creare) un classificatore, la funzione di fitting stima i parametri di una distribuzione gaussiana per ogni classe
* Per stimare le classi di nuovi dati, il classificatore sottoposto a training trova la classe con il costo di classificazione errato più basso

L'analisi discriminante lineare è anche conosciuta come discriminante di Fisher, dal nome del suo inventore.

Nel caso di un problema a due classi ed una funzione discriminante di tipo lineare (LDF), l’obiettivo della fase di addestramento consiste nella stima dei parametri in modo che risulti:



La DF è in questo caso una combinazione lineare delle feature (il parametro è detto bias, mentre il vettore **w** è il vettore dei pesi). L’equazione definisce la regione che separa i punti assegnati a da quelli assegnati a .

Un esempio di un problema a due classi in uno spazio delle feature 2D è riportato in Figura 5 in cui l’obiettivo della fase di learning è quello di trovare i parametri della retta con equazione che separa le due classi.

Mutatis Mutandis è possibile estendere al caso di problema multi-classe e/o con spazio delle feature multidimensionale.

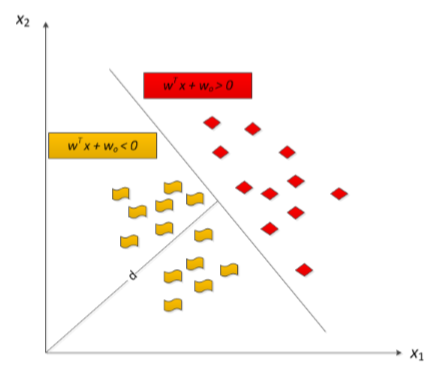


Figura 5: LDF con 2 classi e spazio delle feature 2D

Esistono due varianti principali: l'Analisi Discriminante Lineare (LDA) e l'Analisi Discriminante Quadratica (QDA):

* LDA assume che tutte le classi abbiano la stessa matrice di covarianza, il che porta a funzioni discriminanti lineari.
* QDA permette matrici di covarianza diverse per ogni classe, il che porta a funzioni discriminanti quadratiche.

Oltre a queste varianti sono presenti anche altri tipi di discriminanti come quello lineare con matrici di covarianza diagonali (similarmente al discriminante lineare, ma assume che le matrici di covarianza delle classi siano diagonali: questo significa che le caratteristiche sono considerate indipendenti l'una dall'altra all'interno di ciascuna classe) o quello quadratico con matrici di covarianza diagonali.

**KNN CLASSIFICATION**

Un classificatore molto importante che appartiene alla categoria degli algoritmi non-parametrici è il k-NN (k-Nearest Neighbor) il quale costituisce un algoritmo di ML molto popolare in cui in numero “k” rappresenta il numero di vicini (“Nearest Neighbours”) coinvolti nella decisione in cui il concetto di vicinanza si esprime secondo diverse possibili distanze tra cui quella Euclidea, distanza di Hamming, distanza di Manhattan, distanza di Minkowski ecc.

Un esempio del flusso di funzionamento del classificatore KNN è riportato di seguito (Figura 6) in cui sono presenti due classi (Class A e Class B), ed il vettore caratteristico ha due componenti rappresentabili quindi su un piano x-y: ogni dato corrisponde ad un punto sul piano in cui le stelle rosse appartengono alla classe A e i triangoli verdi alla classe B. L’algoritmo prende la decisione sul nuovo dato (rappresentato dal quadrato giallo con all’interno il punto interrogativo) in base alla “vicinanza” dei K punti più vicini. Nell’esempio K=3 e poiché al nuovo punto che va classificato i 3 punti più vicini appartengono 2 alla classe B e uno alla classe A, l’algoritmo assegna il nuovo punto alla classe B.

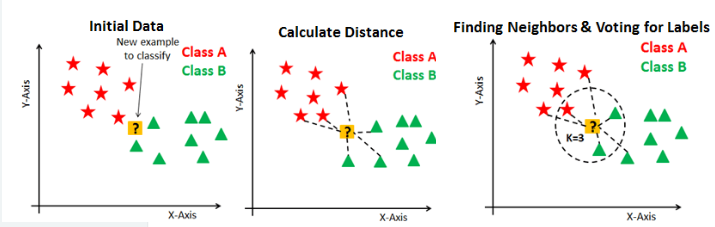


Figura 6: Classificatore KNN

Riassumendo, KNN è un algoritmo non parametrico che classifica i punti dati in base alla loro somiglianza con i punti vicini; esso può essere applicato ad esempio ai dati EEG per identificare i modelli nell'attività cerebrale degli individui con malattia di Parkinson e differenziarli da quelli senza la condizione.

**SVM CLASSIFICATION**

Un classificatore SVM (“Support Vector Machine”) effettua un pre-processing dei dati in modo da rappresentare i pattern in uno spazio di dimensione maggiore di quello originale in modo che i dati possano essere facilmente separabili linearmente da un iperpiano calcolato in modo che il margine tra le varie classi sia massimizzato (Figura 7).

SVM è un algoritmo di apprendimento supervisionato efficace per le attività di classificazione. Esso funziona trovando l'iperpiano ottimale che meglio separa i punti dati in classi diverse. La SVM può essere utilizzata per classificare i segnali EEG in gruppi di controllo sani e della malattia di Parkinson in base alle caratteristiche estratte.

Per questo step occorre selezionare il tipo di kernel come ad esempio lineare, polinomiale, Sigmoid, Gaussian or Radial Basis Function (RBF).

Dopo aver determinato le regioni di decisioni i dati sono proiettati indietro alle dimensioni originali.

Tale classificatore nasce per il caso binario ossia con due classi, ma può essere facilmente adattato al caso multi-classe.

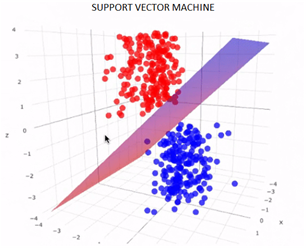


Figura 7: Support Vector Machine (SVM)

L’obiettivo di un algoritmo SVM è quello di trovare un iperpiano che separi, al miglior grado possibile, i punti di dati di una classe da quelli di un’altra classe. Per “migliore” si intende l’iperpiano che ha il margine maggiore tra le due classi, rappresentate dal più e dal meno nella figura qui sotto in Figura 8. I support vector in figura sono i punti di dati nel dataset più vicini alla frontiera di decisione (iperpiano separatore) tra le classi; essi sono fondamentali perché definiscono la posizione e orientamento dell’iperpiano.

Per margine si intende la larghezza massima della linea parallela all’iperpiano che non ha punti di dati interni (in pratica il margine è la distanza dell’iperpiano dai support vector più vicini). L’algoritmo è in grado di trovare un iperpiano di questo tipo solo per i problemi separabili in modo lineare, mentre per i problemi più pratici l’algoritmo massimizza il margine soft, il che consente un numero ridotto di classificazioni errate. Il margine soft è un concetto utilizzato nelle SVM per gestire dati che non sono perfettamente separabili; invece di trovare una separazione perfetta l’algoritmo permette alcuni errori di classificazione, bilanciando tra la massimizzazione del margine e la minimizzazione degli errori di classificazione.

Come detto l’algoritmo SVM standard viene formulato per problemi di classificazione binari e i problemi multi-classe vengono generalmente ridotti a una serie di problemi binari.

Esaminando più approfonditamente l’aspetto matematico, le Support Vector Machine rientrano in una classe di algoritmi di Machine Learning definiti metodi kernel, in cui è possibile trasformare le feature per mezzo di una funzione kernel.

Le funzioni kernel mappano i dati su uno spazio diverso e spesso di dimensione superiore con l’idea che, dopo questa trasformazione, le classi saranno più facili da separare, semplificando potenzialmente i confini decisionali non lineari complessi in confini lineari nello spazio di dimensione superiore delle feature mappate. In questo processo, non è necessario trasformare esplicitamente i dati, il che sarebbe molto dispendioso a livello computazionale: questo processo è tradizionalmente noto come il “kernel trick”.

I kernel rendono le SVM più flessibili e in grado di gestire problemi non lineari.

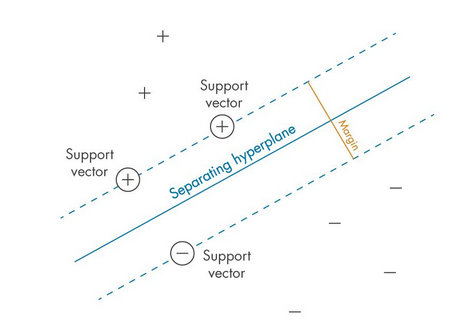


Figura 8: Support Vector Machine (SVM), Margini tra le classi

Le Support Vector Machines sono comunemente utilizzate per problemi di classificazione binaria, ma possono essere estese per affrontare problemi di classificazione multiclasse utilizzando diverse strategie. Ecco alcune delle principali tecniche per estendere le SVM a problemi multiclasse:

* One-vs-All (One-vs-Rest): nella strategia One-vs-All (o One-vs-Rest), si addestrano k classificatori binari, uno per ogni classe, dove è il numero di classi nel problema di classificazione. Ad ogni classificatore viene assegnata una classe di riferimento e l'obiettivo è distinguere quella classe dalle altre. Durante la fase di test, si applicano tutti i classificatori ai nuovi esempi e si seleziona la classe con il massimo valore di decisione.
* One-vs-One: nella strategia One-vs-One, si addestrano ​ classificatori binari, ognuno per distinguere due classi diverse tra le k classi totali. Durante la fase di test, ogni classificatore fornisce un voto per la classe assegnata, e la classe con il maggior numero di voti viene selezionata come previsione finale.

**ANN CLASSIFICATION**

Le Reti Neurali (ANN: Artificial Neural Networks) sono dei modelli costituiti da unità di elaborazione (neuroni) opportunamente connesse dette appunto neuroni, creati in modo da emulare il comportamento dell’encefalo umano. Quando la rete è costituita da molti strati nascosti si parla di Deep Learning (DL) ( Figura 9); si raccomanda per questi tipi di rete il training su GPU anziché su CPU in modo da velocizzare i tempi di elaborazione. Il numero di strati nascosti prende il nome di profondità (“depth”) della rete. Quando la rete non presenta dei loop, quindi i segnali viaggiano in un solo senso dallo strato di input verso quello di output, è detta “feed-forward”; in caso contrario è detta “recurrent”. Il punto cruciale nel progetto della rete è la scelta dei nodi che costituiscono gli strati intermedi, detti Hidden Layers; tale scelta caratterizza il tipo di rete utilizzata (CNN, MLP, RNN, ecc.). Le reti neurali, in particolare i modelli di deep learning come le Reti Neurali Convoluzionali (CNN) o le Reti Neurali Ricorrenti (RNN), possono essere utilizzate per apprendere schemi complessi nei segnali EEG. Questi modelli sono in grado di estrarre automaticamente caratteristiche gerarchiche dai dati e hanno dimostrato potenzialità in varie applicazioni mediche, inclusa la diagnosi di disturbi neurologici.

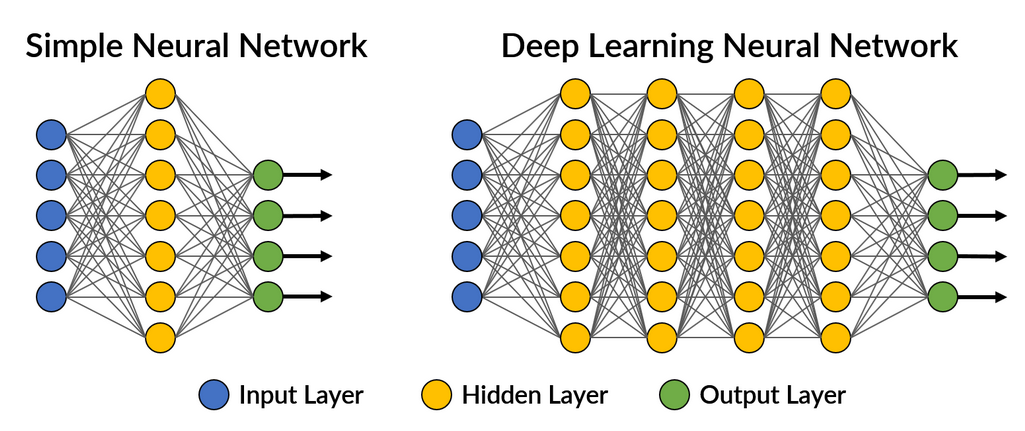


Figura 9: ANN (Artificial Neural Network)

L’addestramento delle Reti Neurali prevedono le seguenti parti:

* Inizializzazione dei Pesi e dei Bias: i pesi e i bias tra i neuroni sono inizializzati casualmente.
* Propagazione in Avanti (“Forward Propagation”): l'input viene passato attraverso la rete neurale, strato per strato, fino allo strato di output. Per ciascun neurone, l'output viene calcolato utilizzando una funzione di attivazione.
* Calcolo della Loss Function: viene calcolata una funzione di loss per valutare la discrepanza tra l'output previsto e il valore reale dell'etichetta di classe.
* Retro-propagazione dell'Errore (“Backpropagation”): l'errore viene propagato all'indietro attraverso la rete, e i pesi e i bias vengono aggiornati utilizzando un algoritmo di ottimizzazione (come la discesa del gradiente) per minimizzare la loss function.
* Iterazione: il processo di propagazione in avanti e retro-propagazione dell'errore viene ripetuto per un numero fisso di epoche o finché la convergenza non è raggiunta.

Il deep learning garantisce un alto livello di precisione, ma richiede grandi quantità di dati per effettuare previsioni accurate (per questo può essere necessario nel caso i dati in input siano immagini effettuare una operazione di “data augmentation” la quale produce un aumento delle immagini del dataset a disposizione a partire dal dataset stesso, apportando delle alterazioni alle immagini come flip, rotazioni, traslazioni, scaling, crop ecc.). Inoltre richiede di avere a disposizione hardware potente, in particolare nella fase di training la quale richiede molto tempo quando la mole di immagine da impiegare è notevole, per questo si consiglia di utilizzare una GPU o un cluster di GPU per velocizzare i tempi.

All’interno del DL, le reti neurali convoluzionali (CNN), o Convolutional Neural Networks, sono un particolare tipo di reti ancora basate sulle caratteristiche (feature-based), ma utilizzano un approccio leggermente diverso rispetto alle reti neurali tradizionali nel processo di estrazione delle caratteristiche. Nelle CNN, le caratteristiche sono estratte direttamente dai dati tramite l'applicazione di filtri convoluzionali su regioni locali dei dati di input. Questi filtri convoluzionali sono progettati per rilevare pattern di basso livello, come linee, bordi e texture, in modo automatico e adattivo, senza la necessità di estrarre manualmente le caratteristiche.

Le reti CNN sono specificamente progettate per l'elaborazione delle immagini e sono diventate uno degli strumenti più potenti e ampiamente utilizzati nell'ambito della visione artificiale e del riconoscimento delle immagini. Se si desidera utilizzare una rete neurale convoluzionale (CNN) per elaborare dati non strutturati che non siano immagini, è comune adottare un approccio che "riconduca" i dati a una forma di input che la CNN possa comprendere. Questo processo può essere chiamato "visualizzazione" o "trasformazione" dei dati in un formato adatto per l'input della CNN: un'opzione comune ad esempio per utilizzare una CNN con segnali audio è quella di convertire il segnale in uno spettrogramma: gli spettrogrammi sono rappresentazioni grafiche che mostrano come la potenza di un segnale varia nel tempo e nella frequenza. Questa rappresentazione consente alla CNN di catturare le caratteristiche spaziali e temporali del segnale audio, rendendola adatta per molti compiti di elaborazione del suono.

Il transfer learning è un approccio molto utile e comune nella classificazione con reti neurali convoluzionali (CNN). In pratica, il transfer learning consiste nel prendere una rete neurale pre-addestrata su un grande dataset e riutilizzarla come punto di partenza per un nuovo task di classificazione. Questo approccio è particolarmente vantaggioso quando si ha a disposizione un dataset relativamente piccolo, poiché permette di sfruttare le conoscenze già acquisite dalla rete pre-addestrata. Uno dei principali vantaggi del transfer learning è proprio il fatto che non è necessario progettare una rete neurale da zero.

Ecco alcuni dei principali vantaggi del transfer learning:

* Riduzione del tempo di sviluppo: utilizzando una rete pre-addestrata, si risparmia tempo che altrimenti sarebbe stato speso nella progettazione e addestramento di una nuova rete da zero. Questo è particolarmente utile quando si lavora su problemi complessi che richiederebbero un lungo periodo di sperimentazione e ottimizzazione.
* Migliori prestazioni con meno dati: le reti pre-addestrate sono state addestrate su grandi dataset e hanno già imparato a riconoscere caratteristiche utili (feature) dalle immagini. Questo significa che possono ottenere buone prestazioni anche su dataset più piccoli, riducendo la necessità di grandi quantità di dati per addestrare una rete da zero.
* Utilizzo di architetture collaudate: i modelli pre-addestrati come VGG, ResNet, Inception (GoogLeNet), AlexNet ecc., sono stati progettati e ottimizzati da esperti e hanno dimostrato di funzionare bene su una vasta gamma di problemi di visione artificiale. Utilizzando queste architetture, ci si può avvalere di design che sono già stati validati nella comunità scientifica.
* Accelerazione dell'addestramento: poiché gran parte della rete è già addestrata, il processo di addestramento per il nuovo task è molto più veloce. Si addestrano solo gli strati finali (e eventualmente si fa fine-tuning), il che richiede meno risorse computazionali e tempo

In genere si rimuove l'ultimo strato della rete pre-addestrata, che è specifico per il task originale in modo da adattarlo al task sotto esame (infatti ad esempio una rete pre-addestrata su 80 classi deve essere modificata per poter identificare ad esempio 5 classi).

**DECISION TREE**

I metodi non-metrici invece che descrivere i pattern con dei vettori di numeri reali, li descrivono in termini di liste di attributi, ad esempio rappresentando i valori di un numero fisso di proprietà in una tupla detta “property D-tuple”. Ad esempio considerando la descrizione di un frutto dalle seguenti quattro proprietà (colore, texture, gusto, dimensione) allora un particolare frutto potrebbe essere rappresentato dalla 4-tuple [rosso, lucido, dolce, piccolo]. Un altro approccio potrebbe essere la descrizione di un pattern da una stringa di lunghezza variabile di attributi, come ad esempio la sequenza di coppie base in un segmento di DNA: “AGCGTTACGATTCCA”. E’ intuitivo quindi classificare un pattern come una sequenza di domande, visualizzata come un albero di decisione (“decision tree”) in cui il primo nodo, visualizzato nella parte alta, prende il nome di radice (“root”), ed è connesso mediante dei link o rami ad altri nodi i quali a loro volta saranno collegati ad altri nodi fino ad arrivare ai nodi terminali detti foglie (“leaf”), le quali non hanno ulteriori rami (Figura 10).

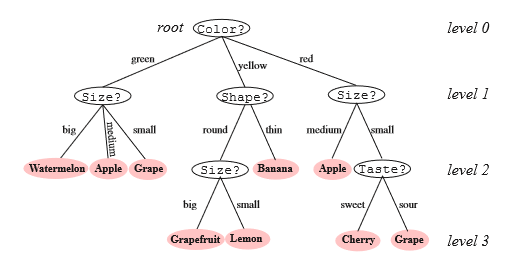


Figura 10: Esempio di Decision Tree

La classificazione di un pattern inizia dal nodo radice, il quale chiede il valore di una particolare proprietà del pattern (i vari rami rappresentano i valori possibili tra cui scegliere). In base alla risposta si segue un dato ramo fino ad arrivare al nodo successivo in cui si ripete il processo. Il meccanismo si ripete fino ad arrivare ad un nodo foglia, il quale non presenta ulteriori domande, bensì produce una label di classe ed il pattern viene assegnato alla categoria del nodo foglia raggiunto. Tra i vari metodi che si possono impiegare per l’induzione degli alberi di decisione si possono citare CART (Classification and Regression Trees), ID3, e C4.5.

Gli alberi sono binari sono molto diffusi: in questo caso ogni passaggio di una previsione comporta il controllo del valore di un predittore (variabile). Un esempio è riportato in Figura 11: questo albero prevede le classificazioni in base a due predittori, x1 e x2 che costituiscono il vettore caratteristico. Per fare una previsione, si inizia dal nodo superiore, rappresentato da un triangolo (Δ). La prima decisione è se x1 è minore di 0.5: in tal caso, si seguire il ramo sinistro e si vede che l'albero classifica i dati come tipo 0. Se, invece, x1 supera 0,5, si segue il ramo destro fino al nodo del triangolo in basso a destra: qui l'albero verifica se x2 è minore di 0,5. In caso affermativo, si segue il ramo sinistro e l'albero classifica i dati come tipo 0; in caso contrario, si segue invece il ramo destro e si può osservare che l'albero classifica i dati come tipo 1.

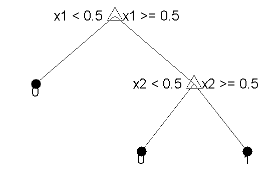


Figura 11: Esempio di Decision Tree Binario

Ricapitolando un albero di decisione è composto da nodi e rami:

* Nodi interni: rappresentano i test sulle caratteristiche.
* Rami: rappresentano i risultati dei test e conducono a ulteriori nodi o foglie.
* Foglie: rappresentano le previsioni finali (classi nel caso di classificazione, valori nel caso di regressione).

Un albero di decisione è considerato binario quando ogni nodo interno ha al massimo due rami (quindi ogni nodo interno divide il dataset in due gruppi basati su una singola caratteristica e una soglia specifica). In altre parole, ogni suddivisione nel processo di costruzione dell'albero crea esattamente due sottoinsiemi dei dati. Questo tipo di albero è spesso usato perché semplifica la struttura dell'albero e la logica di suddivisione.

I passi per la costruzione dell’albero sono i seguenti:

* Si parte dal nodo radice che contiene l'intero dataset.
* Ad ogni passo, l'albero sceglie una caratteristica (feature) e una soglia per suddividere i dati in modo che i sottogruppi risultanti siano quanto più omogenei possibile rispetto alla variabile target.
* Questo processo di suddivisione continua fino a raggiungere un criterio di arresto, come una profondità massima dell'albero, un numero minimo di osservazioni per nodo, o un miglioramento minimo nell'omogeneità.

Durante la costruzione di un albero binario, si utilizzano criteri di suddivisione come l'impurità di Gini, l'entropia o l'errore quadratico medio (MSE) per determinare la migliore caratteristica e soglia per dividere i dati.

I passi per classificare una nuova osservazione sono i seguenti:

* l'osservazione viene fatta passare attraverso l'albero iniziando dal nodo radice.
* in ogni nodo, viene testata la caratteristica corrispondente e si sceglie il ramo appropriato.
* questo processo continua fino a raggiungere una foglia, la quale fornisce la previsione.

Tra le tecniche di ottimizzazione in questo contesto si citano:

* Potatura (“Pruning”): consiste nella rimozione dei rami meno significativi per ridurre il sovra-adattamento o overfitting ai dati di training.
* Random Forest: consiste nella utilizzazione di una collezione di alberi di decisione costruiti su sottocampioni dei dati per migliorare la robustezza e le prestazioni.

Gli alberi decisionali sono modelli semplici ma potenti che possono essere utilizzati per classificare i dati in base a una serie di decisioni gerarchiche. Gli alberi decisionali sono interpretabili e possono fornire informazioni sulle caratteristiche più rilevanti per distinguere tra pazienti con malattia di Parkinson e individui sani.

**RANDOM FOREST**

Il classificatore Random Forest è una tecnica di “ensemble” che può essere utilizzata con alberi di decisione di vari tipi, anche se nella pratica si utilizza comunemente con alberi di decisione binari per semplicità e efficienza. Tuttavia, la tecnica stessa non è limitata agli alberi binari.

Nel nostro caso siamo interessati alla classificazione, non alla regressione quindi l’obiettivo è prevedere la classe o l'etichetta di appartenenza di un'osservazione in base ai suoi attributi quindi l’output di un albero di classificazione è una classe o un'etichetta discreta.

Il Random Forest è un “ensemble learning method”, il che significa che combina i risultati di diversi modelli per migliorare le prestazioni complessive: nel caso specifico del Random Forest, l'ensemble è costituito da una collezione di alberi di decisione.

Nel contesto del Random Forest, il termine "foresta" (forest) si riferisce a una collezione (o un insieme) di alberi di decisione che vengono addestrati in parallelo per risolvere un problema di classificazione o regressione. Ogni albero nella foresta è addestrato su un sottoinsieme casuale dei dati di addestramento tramite il processo di campionamento con sostituzione (bootstrap), e le previsioni di tutti gli alberi vengono combinate per ottenere una previsione finale più accurata e robusta.

Il primo passo è dato dalla creazione della “foresta” da cui appunto la parola “Forest” nel nome del classificatore:

* Il Random Forest crea un insieme di alberi di decisione, di solito centinaia o migliaia.
* Ogni albero è addestrato su un sottoinsieme diverso dei dati di addestramento, utilizzando un processo chiamato bootstrapping (campionamento con sostituzione).

Il processo di bootstrapping (Campionamento Casuale con Sostituzione) prevede:

* Per costruire ogni albero della foresta, si effettua un campionamento casuale con sostituzione dal dataset di addestramento. Questo significa che ogni volta che si estrae un esempio dal dataset per includerlo nel sottoinsieme di addestramento di un albero, si sostituisce l'esempio nel dataset originale, consentendo che lo stesso esempio venga selezionato più volte o che non venga selezionato affatto.
* La dimensione del sottoinsieme di addestramento (detto anche "bag" o "bootstrap sample") è tipicamente pari alla dimensione del dataset di addestramento, ma con alcune istanze potenzialmente duplicate e altre mancanti a causa del processo bootstrap.

Poiché ciascun albero viene addestrato su un diverso sotto-campionamento dei dati, gli alberi saranno diversi tra loro. Questo porta a una maggiore diversità nella foresta, che è essenziale per migliorare le prestazioni complessive.

Per la classificazione, ogni albero nella foresta vota per una classe. La classe con il maggior numero di voti è considerata la previsione finale (per la regressione, le previsioni di ogni albero sono mediate per ottenere la previsione finale). L'aggregazione delle previsioni di tutti gli alberi attraverso il voto o la media tende a ridurre la varianza complessivo del modello, migliorando la capacità di generalizzazione.

Tra i vantaggi dell’ensemble Random Forest si evidenziano i seguenti aspetti:

* Accuratezza: l'ensemble dei molti alberi di decisione tende ad avere prestazioni più accurate rispetto ai singoli alberi.
* Robustezza: riduce il rischio di “overfitting” grazie alla media delle previsioni di molti alberi, ognuno addestrato su diversi sottocampioni e caratteristiche.
* Gestione delle Caratteristiche: Può gestire dataset con molte caratteristiche senza richiedere una selezione preliminare delle stesse.

Il classificatore Random Forest può essere utilizzato per analizzare i dati EEG e identificare caratteristiche importanti per la diagnosi del morbo di Parkinson.