

SCOPO DELL'ESPERIMENTO

L'esperimento si propone di discriminare tra soggetti sani e soggetti affetti da patologie neurodegenerative attraverso l'analisi di campioni di scrittura mediante una rete neurale artificiale di tipo MLP con un livello hidden.

Nel corso dell'esperimento dovranno essere valutati gli effetti dei seguenti parametri sulle prestazioni:

- *dimensione del training set (TS) e del validation set (VS):*

per questo scopo e con riferimento alla dimensione dell'intero dataset, si dovranno considerare dimensioni del TS dal 30% al 50% combinate con dimensioni del VS dal 20% al 50% in modo tale che la dimensione del test set sia in ogni caso non inferiore al 30%.

- *campioni inseriti nel TS e nel VS:*

dovrà essere utilizzata una stratified 5-fold cross validation: per ogni dimensione di TS e VS, dovranno essere scelti in maniera causale i campioni estraendoli dal dataset in modo tale da rispettare la distribuzione dei campioni tra le classi presente nell'intero dataset e tale procedura dovrà essere ripetuta 5 volte.

- *inizializzazione dei pesi;*

per ogni coppia TS-VS estratta come in precedenza dovranno essere eseguiti 10 run, inizializzando in maniera diversa i pesi della rete.

- *learning rate:*

ogni set di 10 run dovrà essere ripetuto con valori del learning rate variabili tra 0,3 e 0,6 con passo da scegliersi opportunamente in funzione di risultati ottenuti.

- *numero di nodi del livello hidden:*

la scelta è libera.

Le prestazioni dovranno essere valutate in termini di *accuracy*, *specificity* e *sensitivity* per ogni coppia VS-TS, mediate sui 10 run. Per la definizione di questi indici di prestazione si può far riferimento alle relative voci di Wikipedia (versione inglese).

Al termine degli esperimenti deve essere prodotto un report con una sintesi dei risultati ottenuti (mediante tabelle o grafici) in grado di evidenziare il ruolo giocato dai parametri di cui sopra sulle prestazioni.

DATASET

Il dataset dal quale sono stati estratti i dati è composto da 136 soggetti dei quali 68 affetti dalla malattia di Alzheimer e 68 che compongono il gruppo dei soggetti sani di controllo. Dei pazienti 32 sono uomini e 36 donne, mentre il gruppo dei soggetti sani composto da 26 uomini e 42 donne. I due gruppi non presentano differenze significative in termini di età e di anni di scolarizzazione.

TASK

I dati sui quali dovranno essere eseguiti gli esperimenti sono ottenuti dall'esecuzione di task di scrittura specifici eseguiti su un foglio di carta poggiato su di una tavoletta grafica che consente di acquisire la sequenza di punti tracciati dalla penna sul foglio. I dati sono stati prodotti da soggetti sani e da soggetti affetti da malattie neurodegenerative ai quali era stato richiesto di eseguire i 3 task descritti nel seguito.

Il primo task richiede di tracciare una circonferenza per quattro volte seguendo una traccia prestampata su un foglio senza mai alzare la penna dal foglio. Il task è pensato per testare l'abilità del soggetto nel compiere movimenti ripetuti e la capacità di coordinazione e pianificazione del movimento. Le capacità di comprensione sono messe alla prova dal momento che è richiesto di tracciare un numero specifico di circonferenze e in maniera continua.

Il secondo task richiede di scrivere il carattere corsivo l ripetuto quattro volte e senza alzare mai la penna dal foglio. Il task ha l'obiettivo di testare le abilità grafo-motorie del soggetto poichè la sua esecuzione richiede il controllo della velocità nel movimento da sinistra verso destra mentre varia di continuo la curvatura del tratto. Come nel task precedente, anche in questo caso è importante mantenere il conto delle ripetizioni e fare attenzione a non alzare la penna dal foglio.

L'ultimo task richiede al soggetto di ricopiare un breve testo. La copia è un processo complesso e verifica non solo la capacità di ripetere una attività ma anche la capacità di organizzazione dello spazio e la capacità di pianificazione di un'attività.

FEATURES

A partire dai dati ottenuti dall'esecuzione di ognuno dei task sono calcolate le features descritte nel seguito.

Nome	Nome nel file CSV	Descrizione
Total Time	(total_time)	Tempo totale impiegato per eseguire l'intero task
Air Time	(air_time)	Tempo impiegato in movimenti in aria durante l'esecuzione del task. Il movimento in aria comprende i movimenti effettuati quando la punta della penna non è in contatto con il foglio di carta.
Paper Time	(paper_time)	Tempo impiegato in movimenti su carta durante l'esecuzione del task. Il movimento su carta comprende i movimenti effettuati quando la punta della penna è in contatto con il foglio di carta.
Mean Speed on-paper	(mean_speed_on_paper)	Media della velocità di movimento su carta della punta della penna durante l'esecuzione del task.
Mean Speed in-air	(mean_speed_in_air)	Media della velocità di movimento in aria della punta della penna durante l'esecuzione del task.
Mean Acceleration on-paper	(mean_acc_on_paper)	Media dell'accelerazione del movimento su carta della punta della penna durante l'esecuzione del task.
Mean Acceleration in-air	(mean_acc_in_air)	Media dell'accelerazione del movimento in aria della punta della penna durante l'esecuzione del task.
Mean Jerk on-paper	(mean_jerk_on_paper)	Media dello strappo del movimento su carta

		della punta della penna durante l'esecuzione del task.
Mean Jerk in-air	(mean_jerk_in_air)	Media dello strappo del movimento in aria della punta della penna durante l'esecuzione del task.
Pressure Mean	(pressure_mean)	Media della pressione esercitata sul foglio durante l'esecuzione dell'intero task.
Pressure Var	(pressure_var)	Varianza della pressione esercitata sul foglio durante l'esecuzione dell'intero task.
Max X Extension	(max_x_extension)	Massima estensione registrata lungo l'asse y durante l'esecuzione dell'intero task. La massima estensione è misurata valutando la differenza tra il punto più lontano dal margine destro del foglio e il punto più vicino allo stesso margine raggiunti durante l'esecuzione del task.
Max Y Extension	(max_y_extension)	Massima estensione registrata lungo l'asse x durante l'esecuzione dell'intero task. La massima estensione è misurata valutando la differenza tra il punto più lontano dal margine superiore del foglio e il punto più vicino allo stesso margine raggiunti durante l'esecuzione del task.
Pendowns Number	(num_of_pendown)	Numero di volte in cui la punta della penna entra in contatto con il foglio.
Size Index	(SI)	L'indice misura la capacità di mantenere una dimensione costante dei caratteri durante la scrittura di un testo.
GMRT on-paper	(gmrt_on_paper)	GMRT calcolato considerando il movimento on paper.
GMRT in-air	(gmrt_in_air)	GMRT calcolato considerando il movimento in air.
Mean GMRT	(mean_gmrt)	Media tra GMRT on-paper e GMRT in-air.
Dispersion Index	(disp_index)	Valuta il tasso di dispersione della traccia di inchiostro sul foglio. Varia tra i valori 0 e 1. 0 nel caso di foglio bianco, 1 nel caso di foglio completamente coperto dalla traccia.
Peaks Number	(peaks_number)	Numero di picchi superiori presenti sulla traccia di inchiostro rilevati nella direzione che va dal margine destro al margine sinistro del foglio
TracingIndex Big Circ	(TI_big_circ)	L'indice misura di quanto i punti della traccia di inchiostro si discostano dalla traccia di riferimento. Più i punti sono vicini alla traccia di riferimento, migliore è l'esecuzione del task e minore è il valore dell'indice
RMSE Big Circ	(MSE_big_circ)	Errore quadratico medio calcolato sull'esecuzione del task 2.

STRUTTURA DEL FILE

I dati sono contenuti nel file. La prima linea del file è una linea di intestazione che contiene il nome delle features presenti.

Le restanti 136 linee (una per ogni soggetto) contengono le features descritte estratte da ognuno dei 3 task (per un totale di 66 valori) e la classe di appartenenza del soggetto che ha eseguito il task: A per i soggetti sani, P per quelli affetti dalla patologia.