

ANALISI DEI DSA CON TECNICHE DI MACHINE LEARNING

ABSTRACT

In questa tesi, viene illustrata una ricerca che ha coinvolto 40 ragazzi di scuola secondaria di secondo grado a cui è stata fatta una diagnosi di DSA. I dati dei test somministrati sono stati messi in relazione tra loro e analizzati con le moderne tecniche di machine learning.

Abbiamo individuato una correlazione diretta tra l'Indice di Ragionamento Percettivo IRP e il Trog-2.

Relativamente alla lettura, nei soggetti dislessici del nostro campione è presente un deficit più evidente (e duraturo) nella rapidità di lettura, se confrontato con la correttezza.

Relativamente alle prove di scrittura, svolte in condizione normale e in condizione di soppressione articolatoria, emerge che le prestazioni di scrittura peggiorano notevolmente se svolte in condizione di soppressione articolatoria, sia per l'aspetto correttezza (dettato di parole) sia per l'aspetto velocità (scrittura di numeri in lettere).

Sono emerse correlazioni significative tra la memoria di lavoro (IML), analizzata tramite WISC IV e WAIS IV e tutti gli indici della BDE-2, in particolare con l'area del calcolo.

L'indice della memoria di Lavoro IML risulta essere ben correlato anche con lo span verbale numerico in modo diretto.

Il risultato più interessante ottenuto dalle nostre analisi è certamente la realizzazione di un modello "black box" di memoria di lavoro secondo cui, avendo a disposizione solo l'età del soggetto, lo span di memoria verbale e il Quoziente del Numero della BDE-2 saremmo in grado di stimare anche il valore dell'Indice di Memoria di lavoro di WISC IV e WAIS IV, senza somministrarne le relative prove, con un'accuratezza del 83%.

Modelli di questo tipo gettano le basi per un nuovo scenario futuro di valutazione neuropsicologica dei DSA: se avessimo, infatti, modelli allenati su un grande numero di soggetti di età diverse, sarebbe possibile, durante un percorso di valutazione, somministrare solo i test o addirittura solo gli items che risultano essere dal modello più predittivi dello specifico disturbo. Tale scenario è tutt'altro che utopistico, in quanto tale modalità viene già applicata ogni giorno in un'infinità di ambiti (ad esempio la pubblicità mirata che viene fatta in base alle preferenze degli utenti altro non è che un modello che utilizza dati precedentemente acquisiti).

Dott.ssa Marianna Vallone
psicologa e psicoterapeuta
Contatti: cell: 3518871632
Email: info@sostegnopsicologico.net



UNIVERSITÀ
DEGLI STUDI
DELLA REPUBBLICA
DI SAN MARINO



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

MASTER UNIVERSITARIO DI II LIVELLO IN
Valutazione e Trattamento Neuropsicologico dei
Disturbi Specifici di Apprendimento
VIII edizione

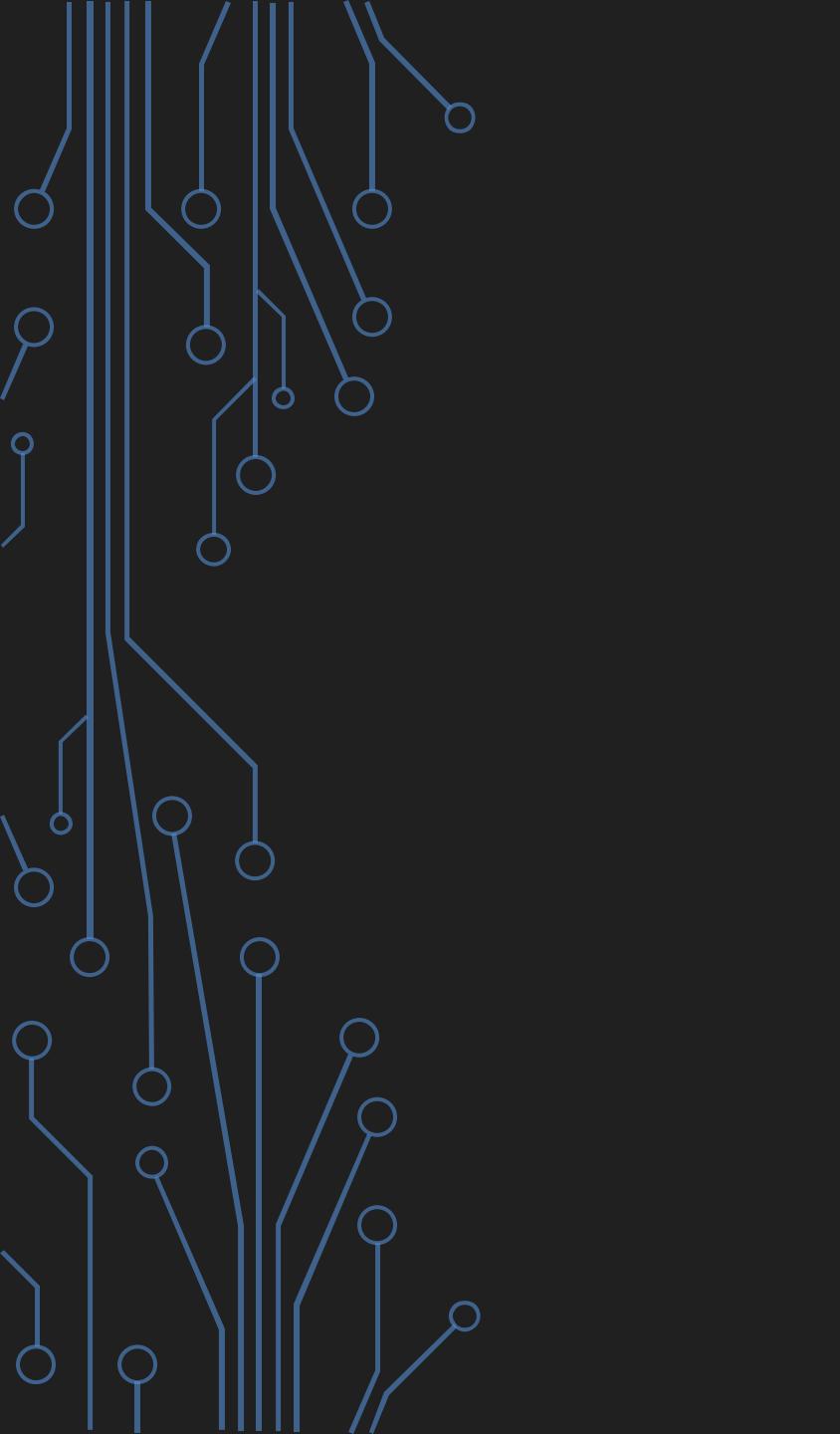
Direttore: Prof. Giacomo Stella

Tesi a cura di:
Marianna Vallone

Relatore:
Maristella Scorza

Titolo tesi:

**ANALISI DEI DSA
CON TECNICHE DI
MACHINE LEARNING**



**IN QUESTA TESI, VERRÀ ILLUSTRATA
UNA RICERCA SUI DISTURBI
SPECIFICI DELL'APPRENDIMENTO
(DSA),
I CUI DATI SONO STATI ANALIZZATI
CON LE MODERNE TECNICHE DI
MACHINE LEARNING.**

OBIETTIVI DELLA RICERCA



indagare le caratteristiche dei DSA nei soggetti di scuola secondaria di secondo grado



ricercare le relazioni tra i dati raccolti



utilizzare le tecniche di machine learning, in particolare le reti neurali artificiali, per confermare le relazioni trovate e per provare a creare dei modelli di previsione

IPOTESI DI RICERCA

che esistano relazioni tra i dati raccolti con i differenti test

che tali relazioni possano essere indagate con le tecniche di machine learning

che si possano creare dei modelli di previsione affidabili, che riescano a predire il funzionamento di un determinato fenomeno

CAMPIONE CLINICO

Ragazzi che hanno richiesto una valutazione neuropsicologica presso il centro SOS Dislessia di Modena.



sono stati selezionati 40 soggetti

**maschi e
femmine**

**frequentanti tutti
la scuola
secondaria di
secondo grado**

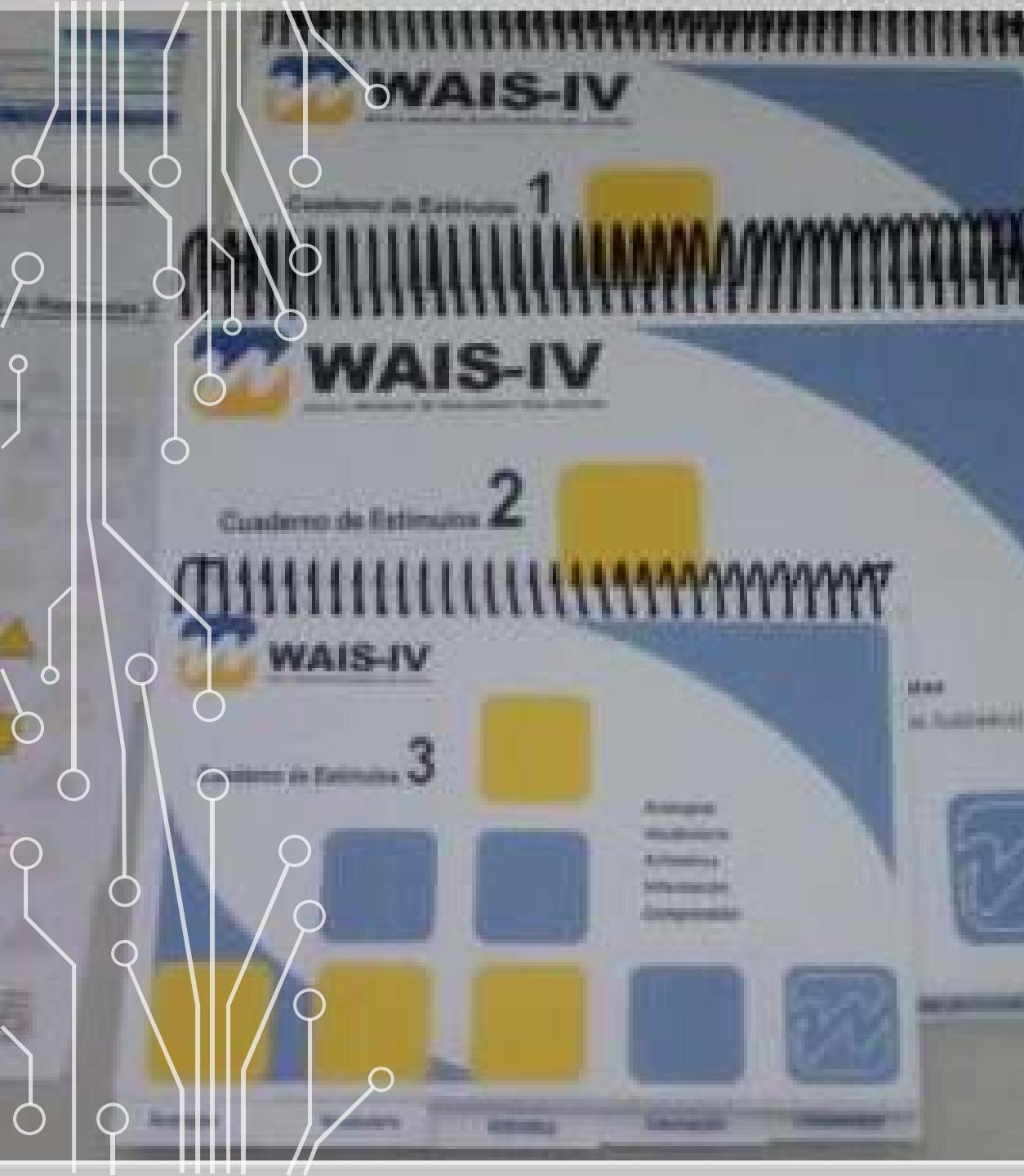
**a cui è stata effettuata,
alla fine della
valutazione,
una diagnosi di DSA**

STRUMENTI UTILIZZATI: IL MACHINE LEARNING

Sono tecniche di analisi dei dati che hanno il vantaggio di **riuscire a lavorare contemporaneamente su un grandissimo numero di dati e di variabili**, e che permettono di **creare dei modelli di descrizione** dei fenomeni, basati sulle reti neurali biologiche.



Le tecniche di machine learning, comportandosi in modo simile alle reti neurali, apprendono le connessioni tra le informazioni, al punto tale da riuscire a fare poi una previsione del fenomeno stesso che ne è alla base.



FASI DELLA RICERCA

1. VALUTAZIONE NEUROPSICOLOGICA:

Sono stati somministrati test di **valutazione della lettura, della scrittura, del calcolo**; test di valutazione degli **indici cognitivi**; test di valutazione della **memoria a breve termine di tipo verbale, numerico e spaziale**; test di valutazione del **linguaggio recettivo ed espressivo**.

FASI DELLA RICERCA

2. IMPORTAZIONE DEI DATI:

In questa fase **sono stati estratti i dati** dal database.

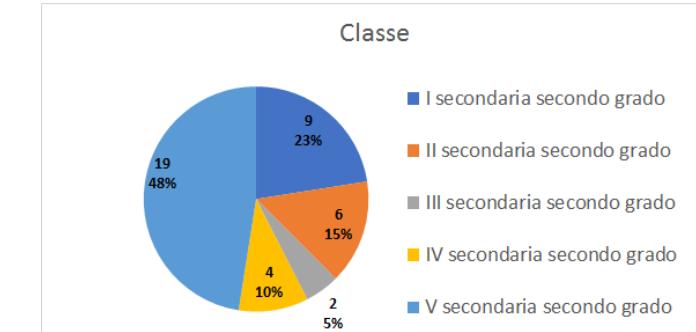
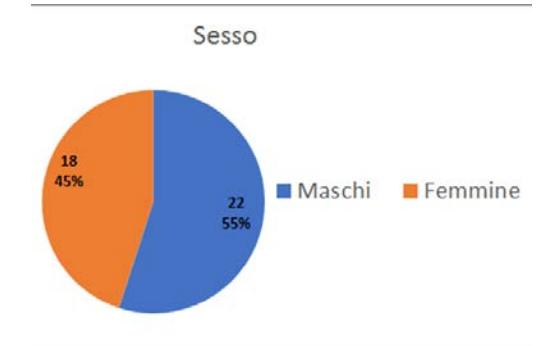
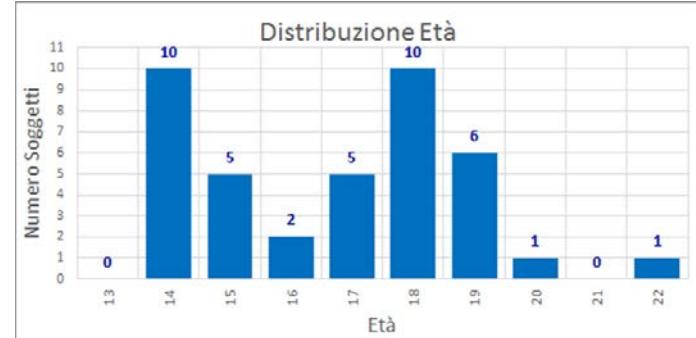
Abbiamo organizzato i dati in **una tabella**, in formato Excel: ad ogni riga corrisponde un'osservazione (un paziente) e ad ogni colonna corrisponde una variabile acquisita (punteggi dei singoli test, dati dell'anamnesi familiare...)

Idx	Datadiosservazione	Datadinnascita	Eta-Mesi	Sesso	Classe	Diagnosi_ICD_10_Texto	Diagnosi_ICD_10_Codice	letturaDiParole_TipoDiTest	LetturaDiParole_tempo_grezzo	LetturaDiParole_tempo_Zscore	LetturaDiParole_velocita_Sil_Sec	LetturaDiParole_velocita_Zscore
29	24/10/2019	18/10/2001	0	F	IV secondaria secondo grado	Discalculia	F81.2	DDE2	54	-0,5	5,2	-0,62
30	31/10/2019	10/2/2005	8	M	I secondaria secondo grado	Disgrafia	F81.8	DDE2	61	0,0	4,61	-0,17

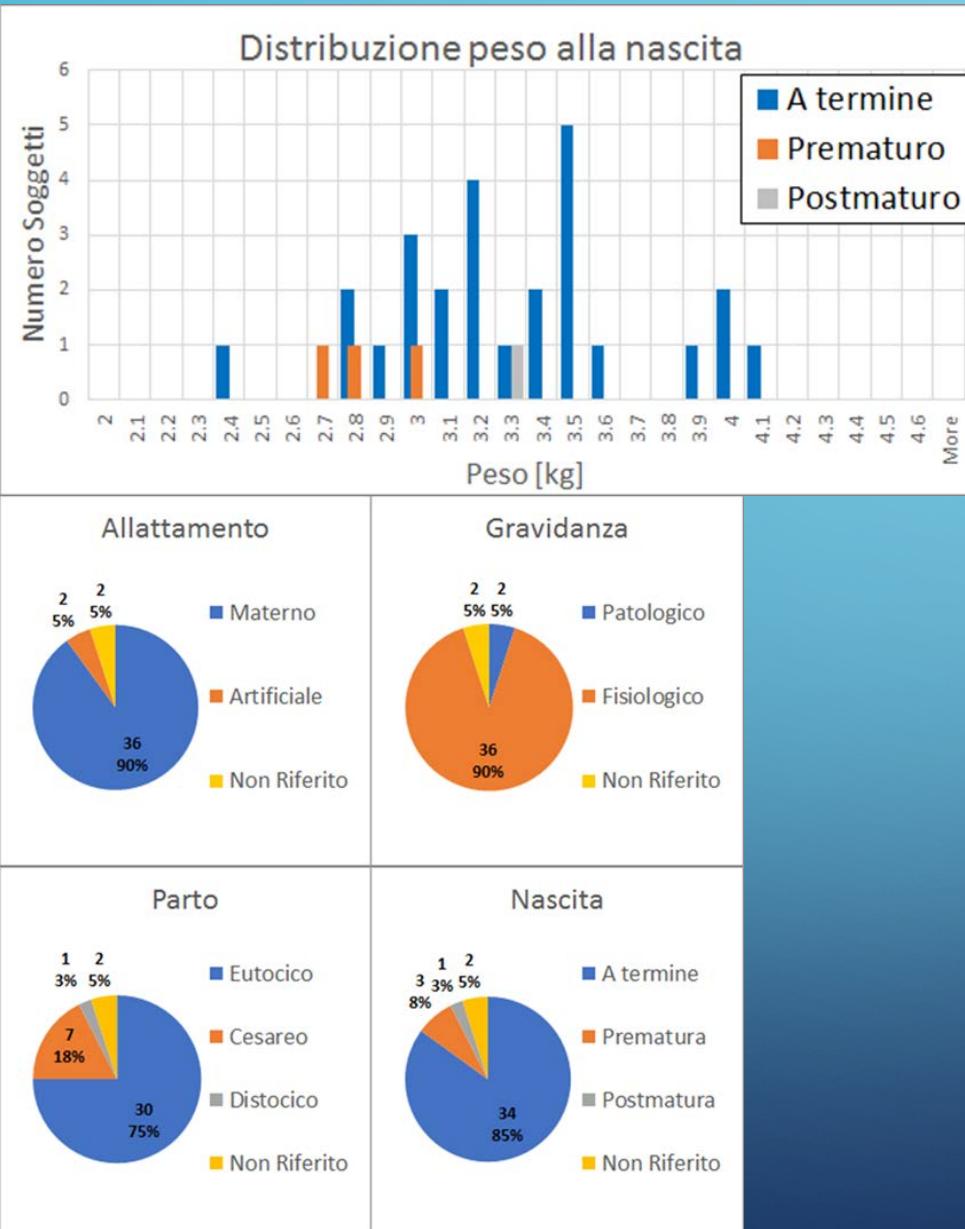
FASI DELLA RICERCA

3. ANALISI DI DATI: DESCRIZIONE DEL CAMPIONE

In questa fase, è stata eseguita una descrizione macroscopica del campione, attraverso il **calcolo delle percentuali e delle distribuzioni**.



3. ANALISI DEI DATI: DESCRIZIONE DEL CAMPIONE



	TOT	%
Lettura		
Lettura di parole		
MT_16_19	17	43%
MT_3	3	8%
DDE2	20	50%
Lettura di non parole		
MT_16_19	17	43%
MT_3	3	8%
DDE2	20	50%
Lettura di Brano		
MT_16_19	20	50%
MT	5	13%
MT_3	15	38%
Comprensione		
MT_16_19	13	33%
MT_3	10	25%
MT	14	35%

	TOT	%
Scrittura		
Frasi Omofone		
DDE2	12	30%
MT_16_19	15	38%
Numeri in lettere		
MT_16_19	19	48%
MT_3	2	5%
BVSCO - Prova 4c	13	33%
Dettato di parole		
MT_16_19	19	48%
MT_3	3	8%
Dettato di brano		
BVSO-Prova1	18	45%
MT_3	9	23%

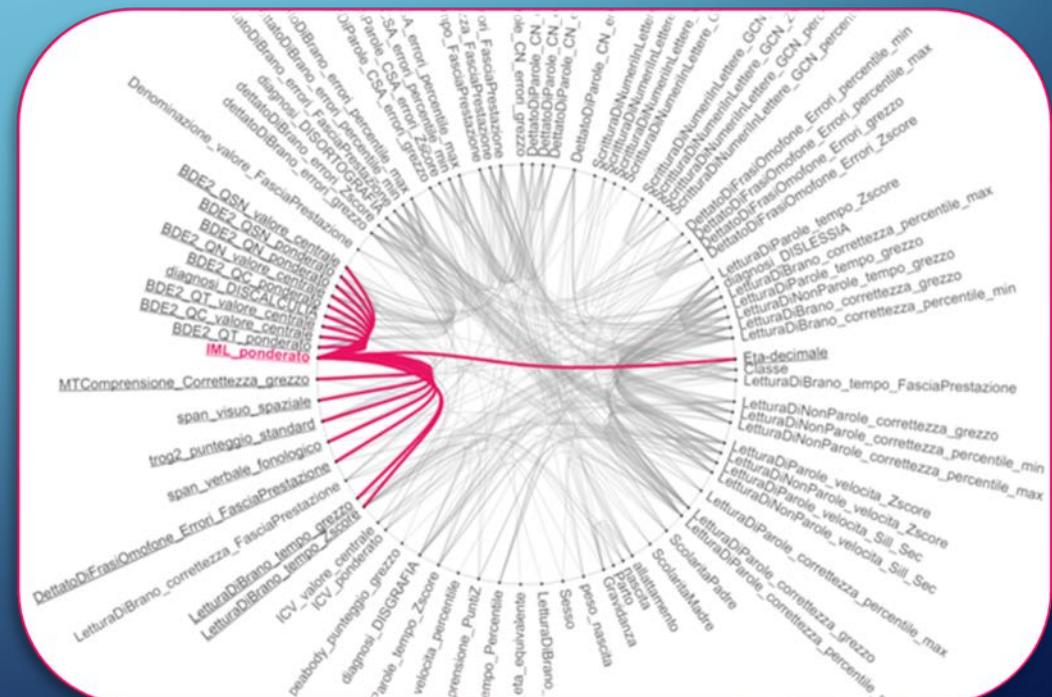
	TOT	%
COGNITIVO E MATEMATICA		
Cognitivo		
WISC-IV	17	43%
WAIS-IV	23	58%
Matematica		
BDE2	40	100%

	TOT	%
ALTRI TEST		
Linguaggio Espressivo		
Denominazione Sartori	33	83%
Boston Naming Test	7	18%
Linguaggio Recettivo		
Peabody	40	100%
Trog	40	100%
Memoria-Span Verbale		
Memoria Breve Termine Verbale	40	100%
Memoria-Span Visuo Spaziale		
Test Corsi	40	100%
Memoria-Span Verbale Numerico		
Memoria Breve Termine Numerica	40	100%
Attenzione Visiva		
BIA CP	40	100%

FASI DELLA RICERCA

4. ANALISI DEI DATI: ANALISI MACROSCOPICA DELLE CORRELAZIONI

- Abbiamo calcolato le correlazioni tra tutte le variabili del campione inserite da noi in tabella utilizzando **MatLab**.
 - Abbiamo così ottenuto **una lista di correlazioni**.
 - Sono stati poi creati dei **grafici a corda** per facilitare l'esplorazione delle correlazioni e per ragionare sulle più significative.



FASI DELLA RICERCA

5. ANALISI DEI DATI: ANALISI SPECIFICA DELLE CORRELAZIONI

Sono state analizzate le correlazioni emerse, escludendo le correlazioni più basse ed anche quelle spurie.

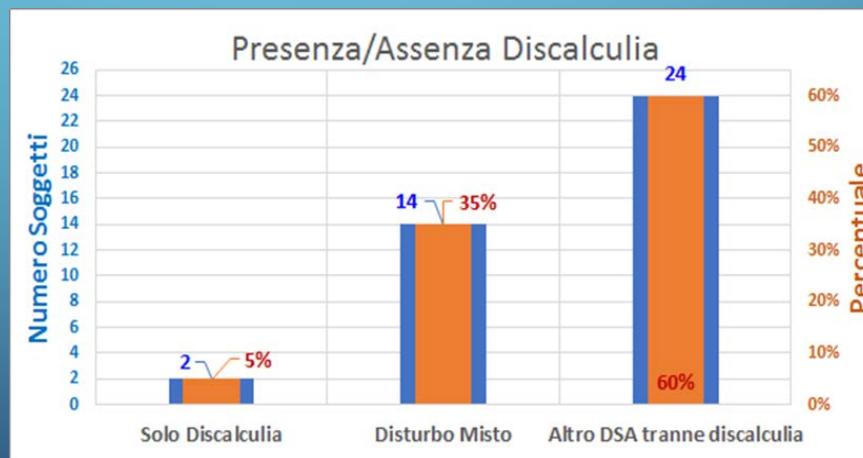
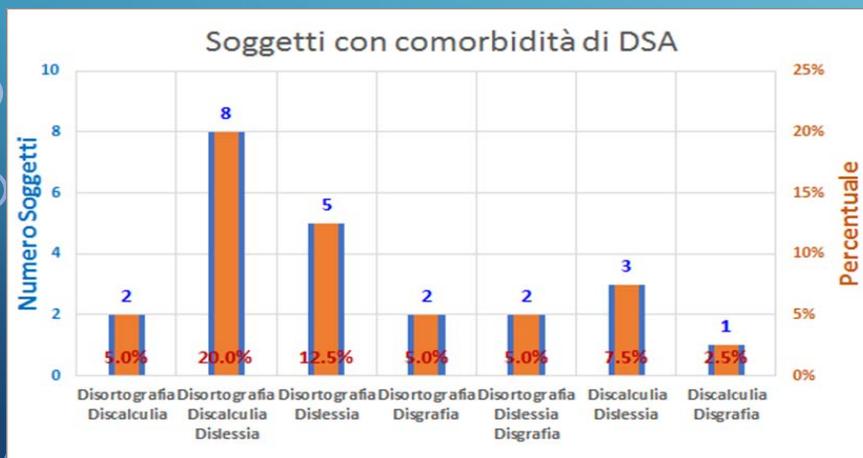
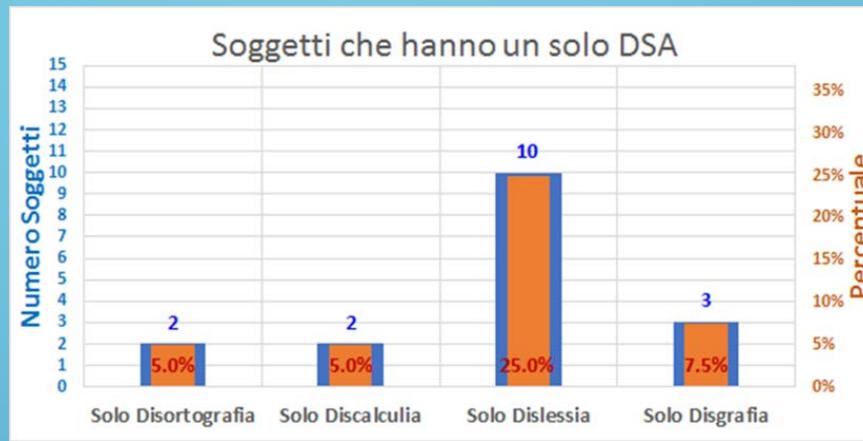
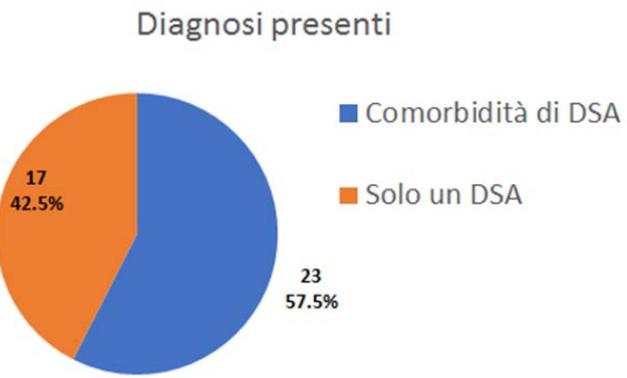
Sono state identificate le correlazioni più interessanti, basandoci sia sul ragionamento clinico, sia sulla letteratura esistente.

Per ogni correlazione **è stata eseguita un'analisi più dettagliata**, attraverso l'uso di diagrammi di dispersione.

6. ANALISI DEI DATI: MODELLAZIONE

Sono state, infine, utilizzate le tecniche di machine learning, per creare un **modello di tipo “black box”**, con la tecnica delle reti neurali.

RISULTATI



DSA isolato: 42,5 %
DSA in comorbidità: 57,5%
DE isolato: 25%
DE in comorbidità: 45%
Discalculia isolata: 5%
Disturbo misto: 35%

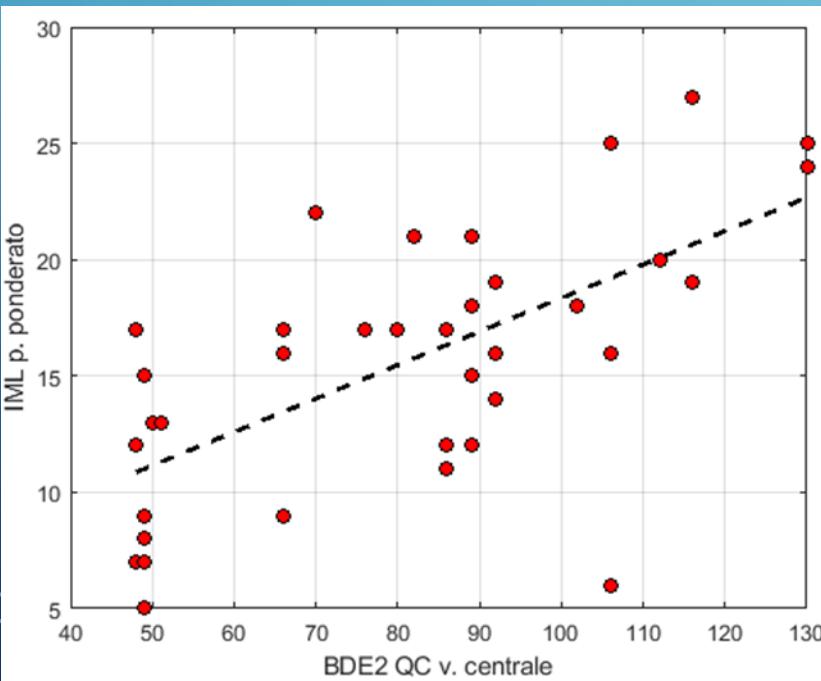
RISULTATI

I dati ci mostrano punteggi più bassi in IML e IVE, rispetto a ICV, IRP e a QIV.

	VALORI STATISTICI		
	<i>M</i>	<i>SD</i>	<i>N</i>
ICV - Valore Centrale	105.8	15.7	16
IML - Valore Centrale	87.3	16.4	40
IRP - Valore Centrale	106.1	13.5	37
IVE - Valore Centrale	93.9	12.0	40
QIV Peabody	116.6	10.0	40
Span Verbale	4.2	0.7	40
Span Visuo Spaziale	5.4	1.2	40
Span Verbale Numerico	5.1	1.0	40

L'INDICE DI MEMORIA DI LAVORO (VARIABILE DIPENDENTE) RISULTA CORRELATO CON TUTTI GLI INDICI DELLA BDE-2:

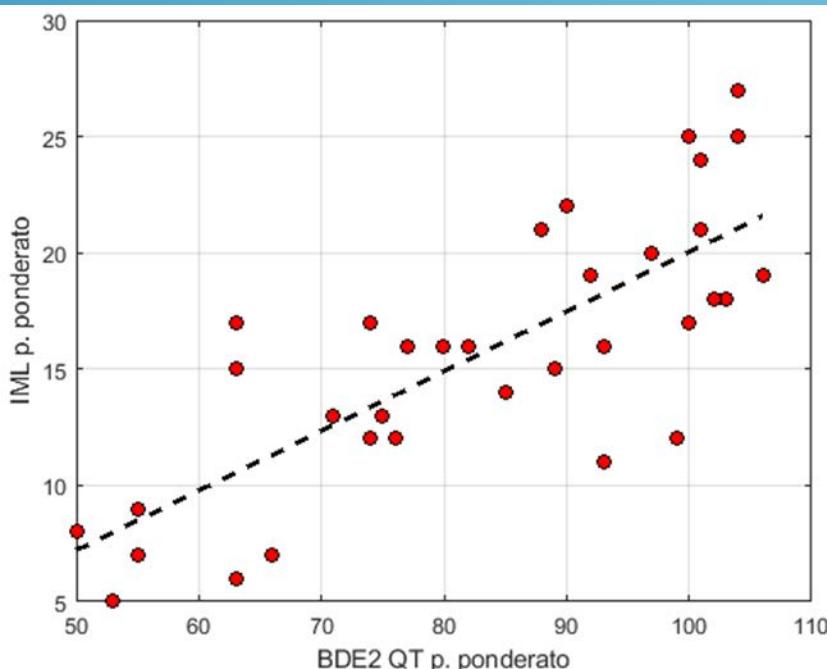
1. L' IML corrella in modo diretto con il Quoziente di Calcolo della BDE-2 (valore centrale e punteggio ponderato), con un coefficiente r pari a .65



Il diagramma di dispersione mostra che: all'aumentare del p.p. della Memoria di Lavoro (WISC IV e WAIS IV) aumenta anche il valore centrale del Quoziente di Calcolo della BDE-2.

L'INDICE DI MEMORIA DI LAVORO (VARIABILE DIPENDENTE) RISULTA CORRELATO CON TUTTI GLI INDICI DELLA BDE-2:

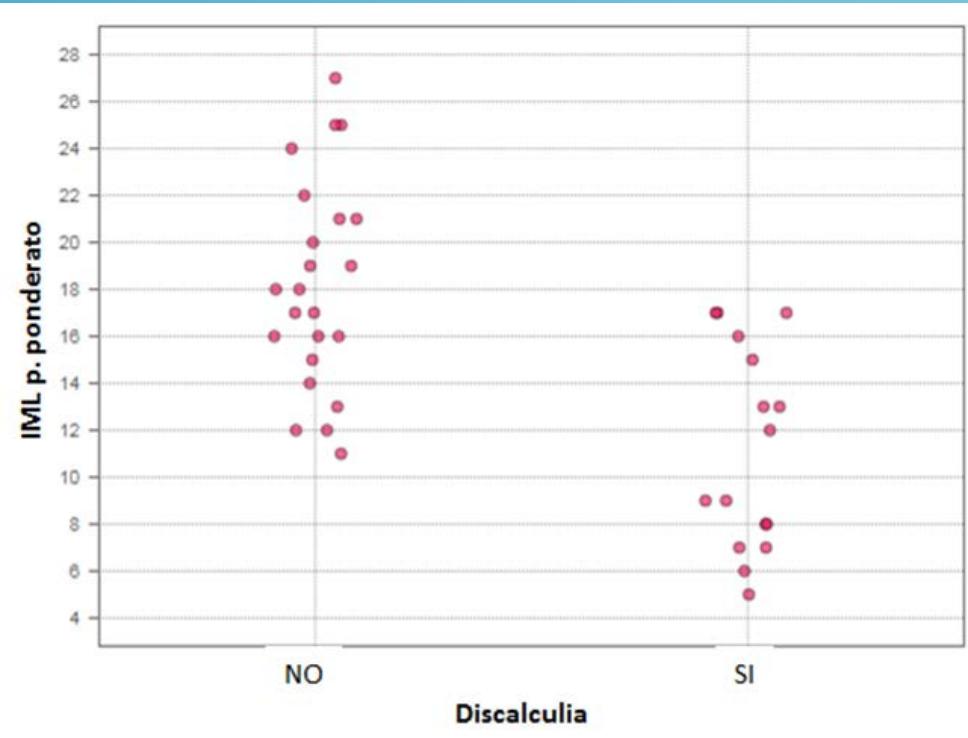
2. L' IML correla in modo diretto con il Quoziente Totale della BDE-2 (valore centrale e punteggio ponderato), con un coefficiente r pari a circa .58.



Il diagramma di dispersione mostra che: all'aumentare del p.p. della Memoria di Lavoro (WISC IV e WAIS IV), aumenta anche il p.p. del Quoziente Totale della BDE-2.

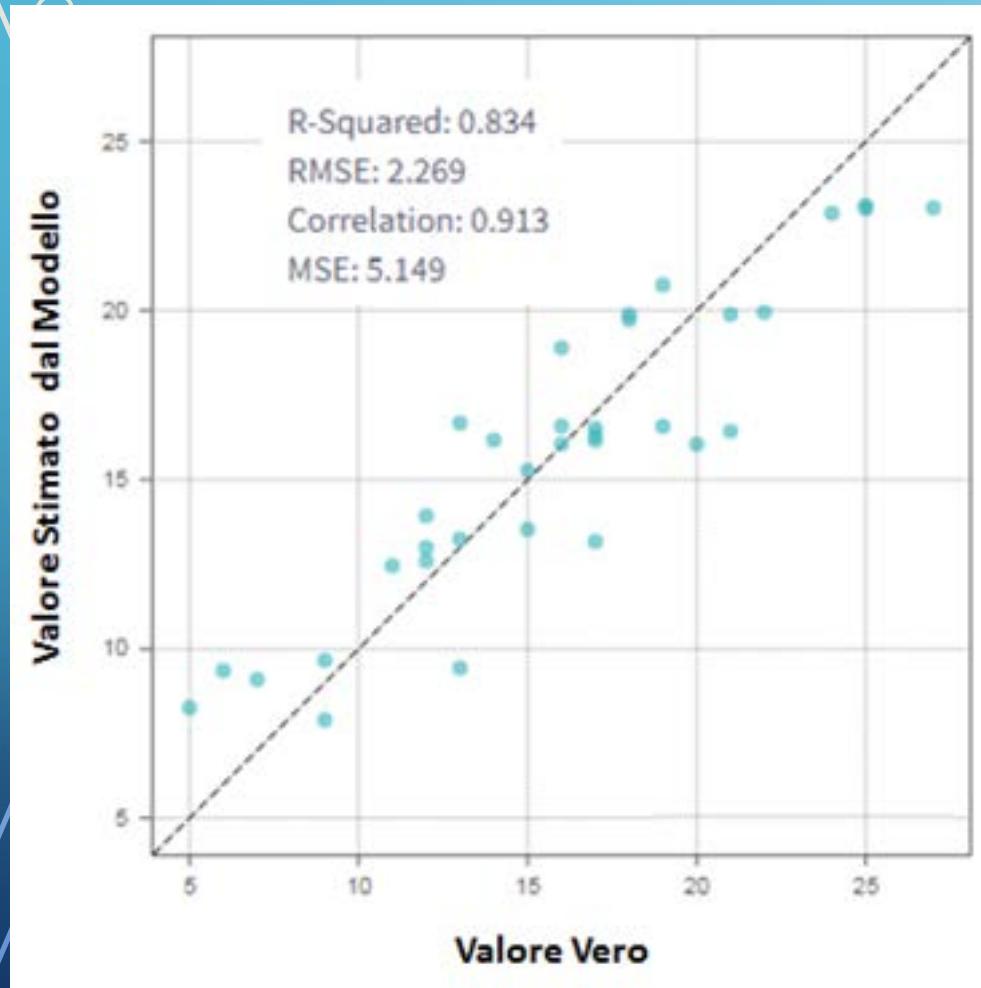
L'INDICE DI MEMORIA DI LAVORO (VARIABILE DIPENDENTE) RISULTA CORRELATO CON TUTTI GLI INDICI DELLA BDE-2:

3. L'IML correla con la diagnosi di discalculia



Il diagramma di dispersione mostra che: i soggetti con diagnosi di discalculia mostrano punteggi ponderati di IML più bassi nella WISC IV e nella WAIS IV.

MODELLO BLACK BOX

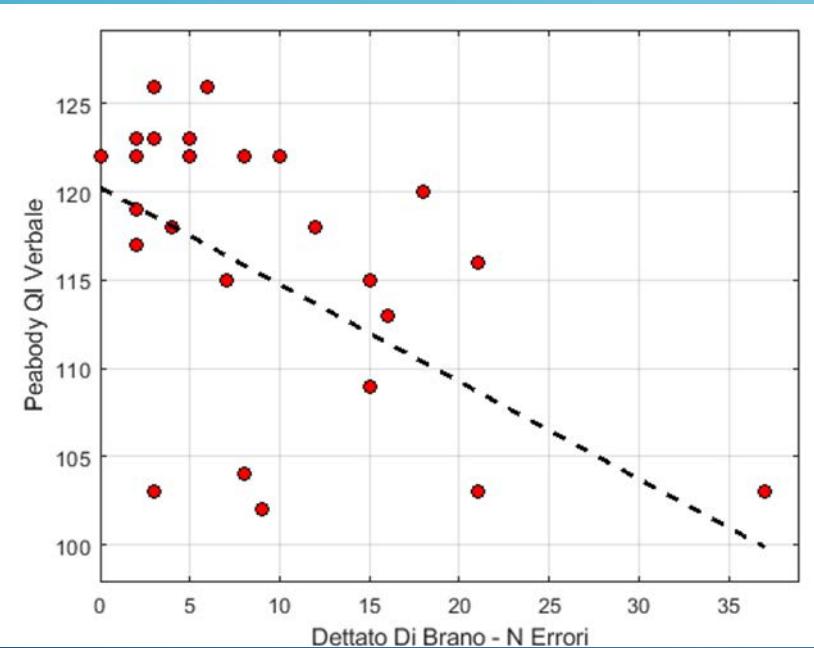


Abbiamo applicato le tecniche di **machine learning** per creare, con MatLab dei **modelli previsione**.

Sono stati allenati migliaia di modelli con i nostri dati e il piu' interessante è risultato essere il modello di previsione dell'Indice di Memoria di Lavoro della WISC IV e della WAIS IV a partire da:

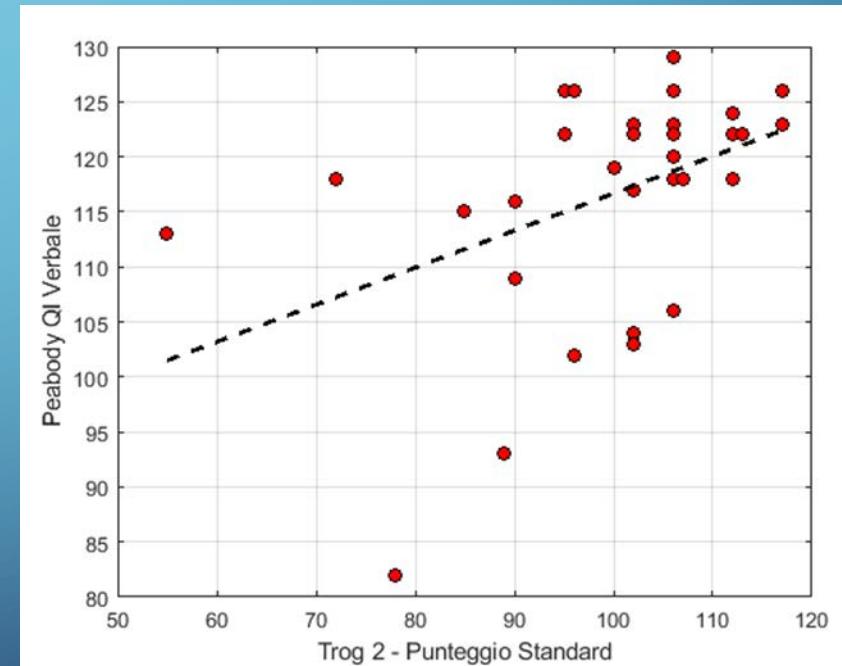
- **età del soggetto**
- **punteggio di span verbale**
- **QN (p.p.) della BDE-2.**

CORRELAZIONE TRA QIV (PEABODY) E IL NUMERO DI ERRORI NEL DETTATO DI BRANO



La correlazione è inversa: all'aumentare del numero di errori nel dettato di brano, diminuisce il QIV.

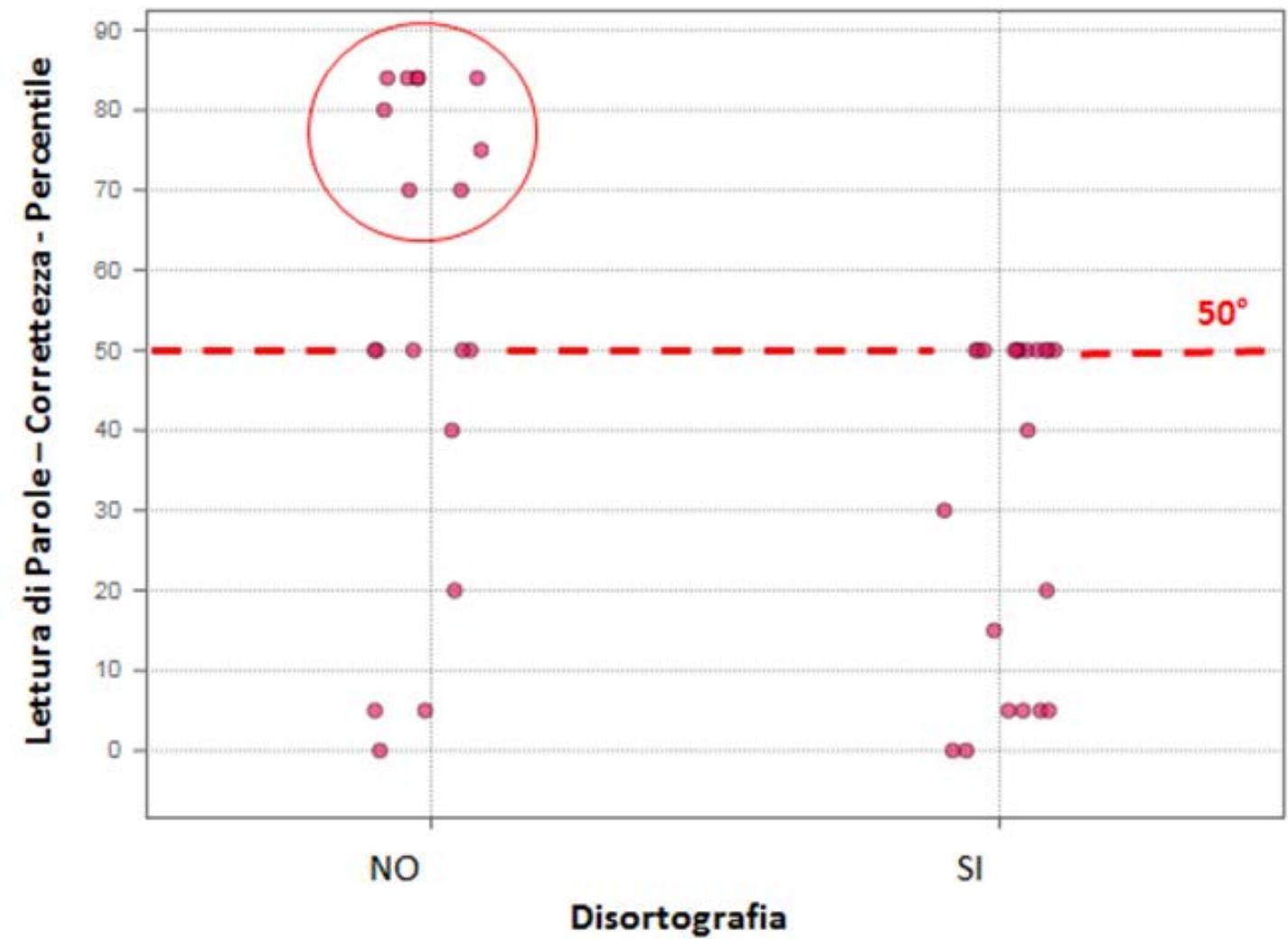
CORRELAZIONE TRA QIV (PEABODY) PEABODY E TROG-2



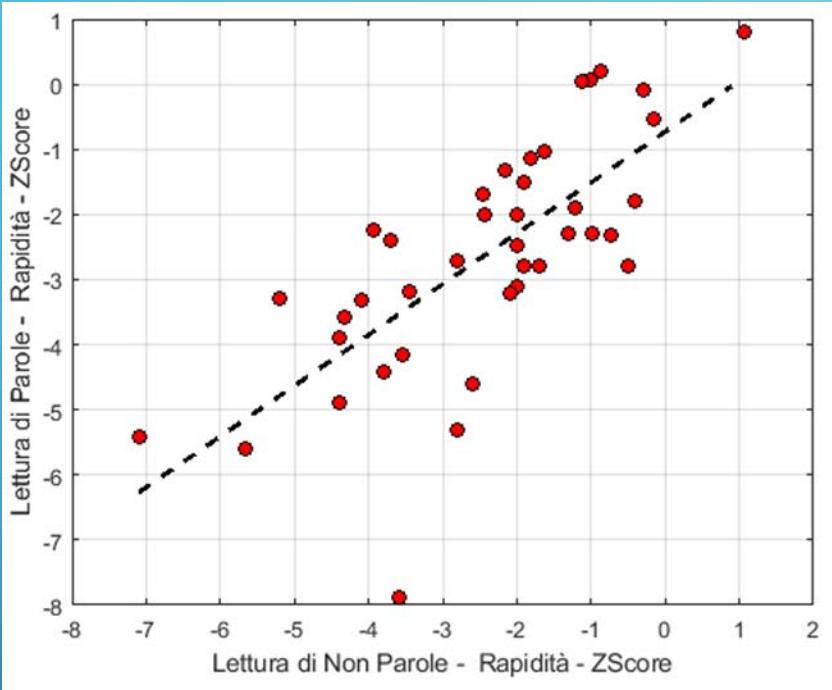
La correlazione è diretta: all'aumentare del QIV del Peabody aumenta anche il punteggio standard del Trog-2.

LETTURA DI PAROLE: CORRETTEZZA

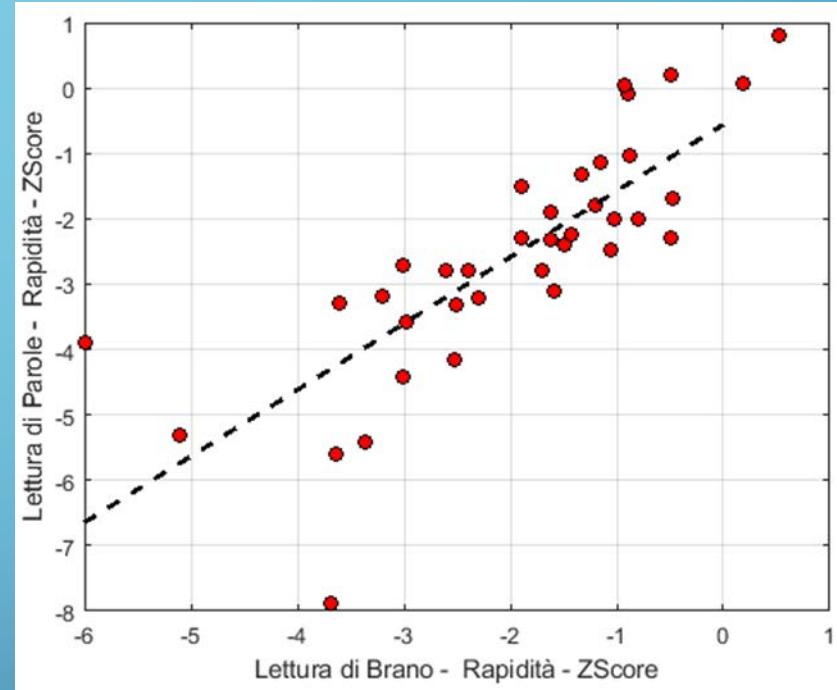
- I soggetti che rientrano in un percentile alto nella lettura di parole non presentano diagnosi di disortografia.
- I soggetti, invece, con diagnosi di disortografia rientrano tutti in percentili pari o inferiori a 50°



LETTURA DI PAROLE: RAPIDITA'



Vi è una correlazione molto alta (coefficiente r pari a .73) tra il tempo impiegato nella lettura di parole e quello impiegato nella lettura di non parole.

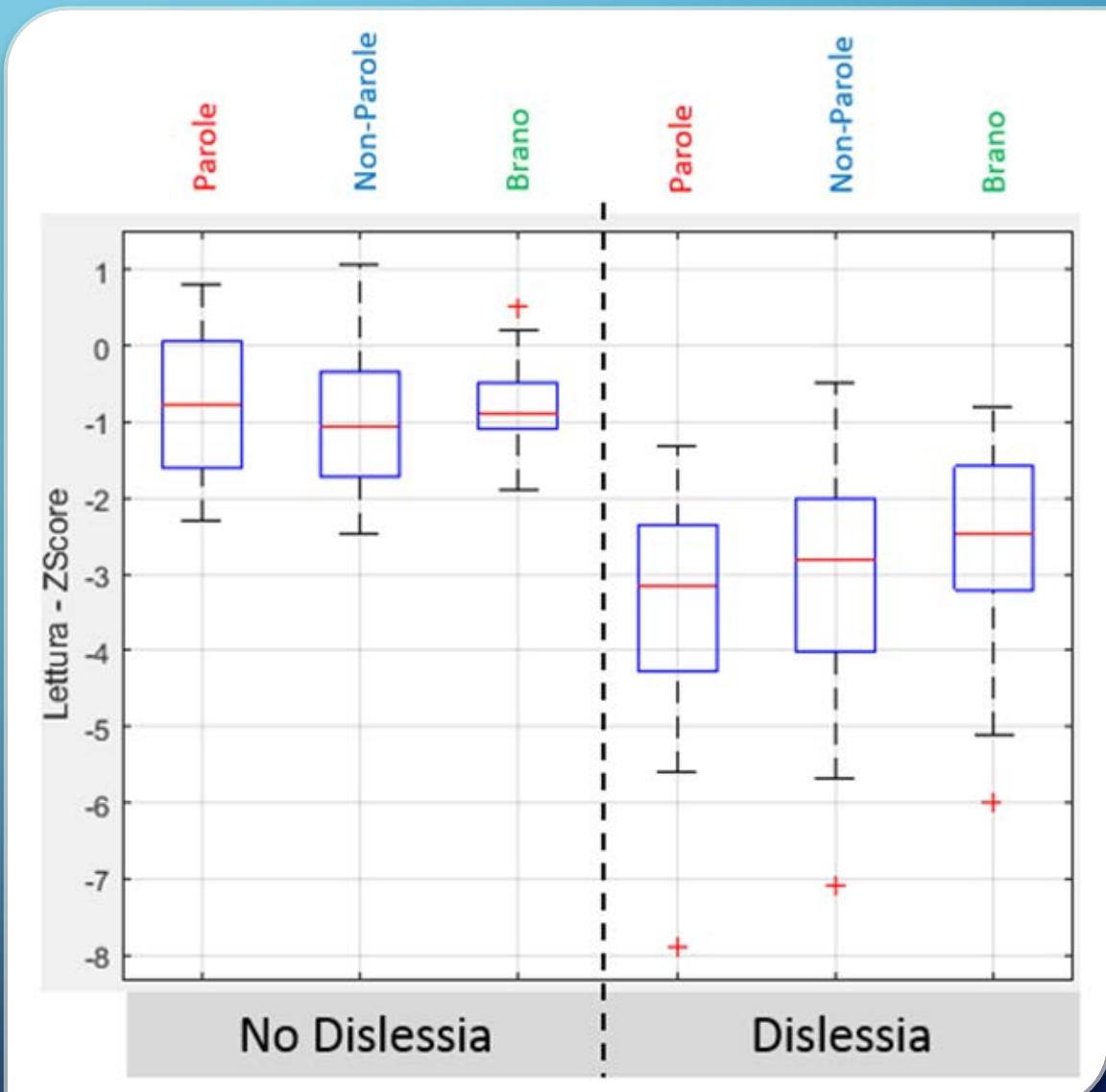


Vi è una correlazione molto alta (coefficiente r pari a .79) tra il tempo impiegato nella lettura di parole e quello impiegato nella lettura di brano.

CARATTERISTICHE DELLA LETTURA NEI SOGGETTI CON DISLESSIA

I soggetti con dislessia, relativamente alla **rapidita' di lettura**, sono molto più lenti dei soggetti senza dislessia ma con altro DSA, in tutti i tipi di lettura.

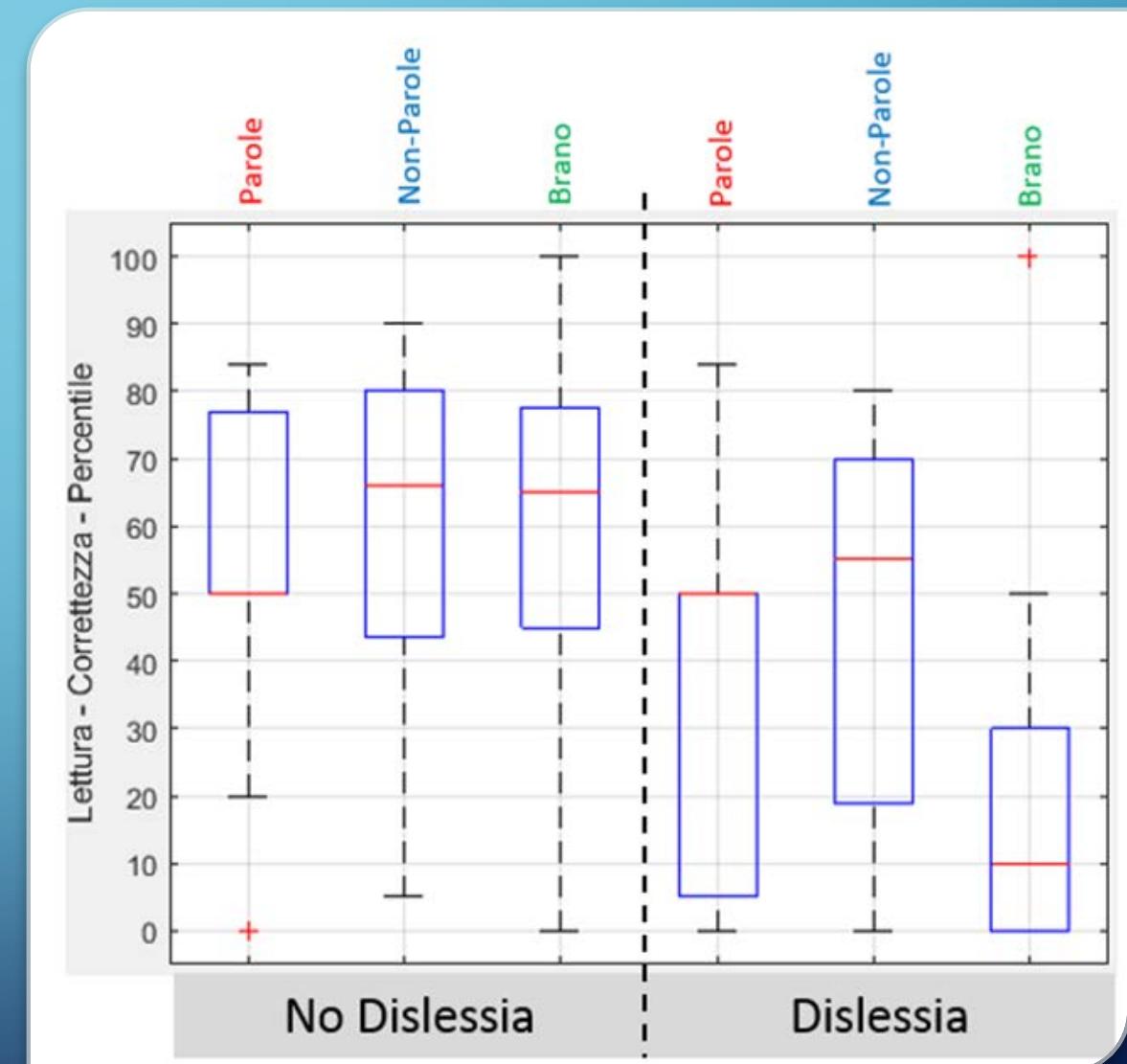
Questo vale, in particolar modo, per la lettura di parole, segue la lettura di non parole e infine la lettura di brano.



CARATTERISTICHE DELLA LETTURA NEI SOGGETTI CON DISLESSIA

I soggetti con dislessia, relativamente alla **correttezza di lettura**, si trovano in percentili più bassi rispetto ai soggetti senza dislessia ma principalmente sopra il 10° percentile.

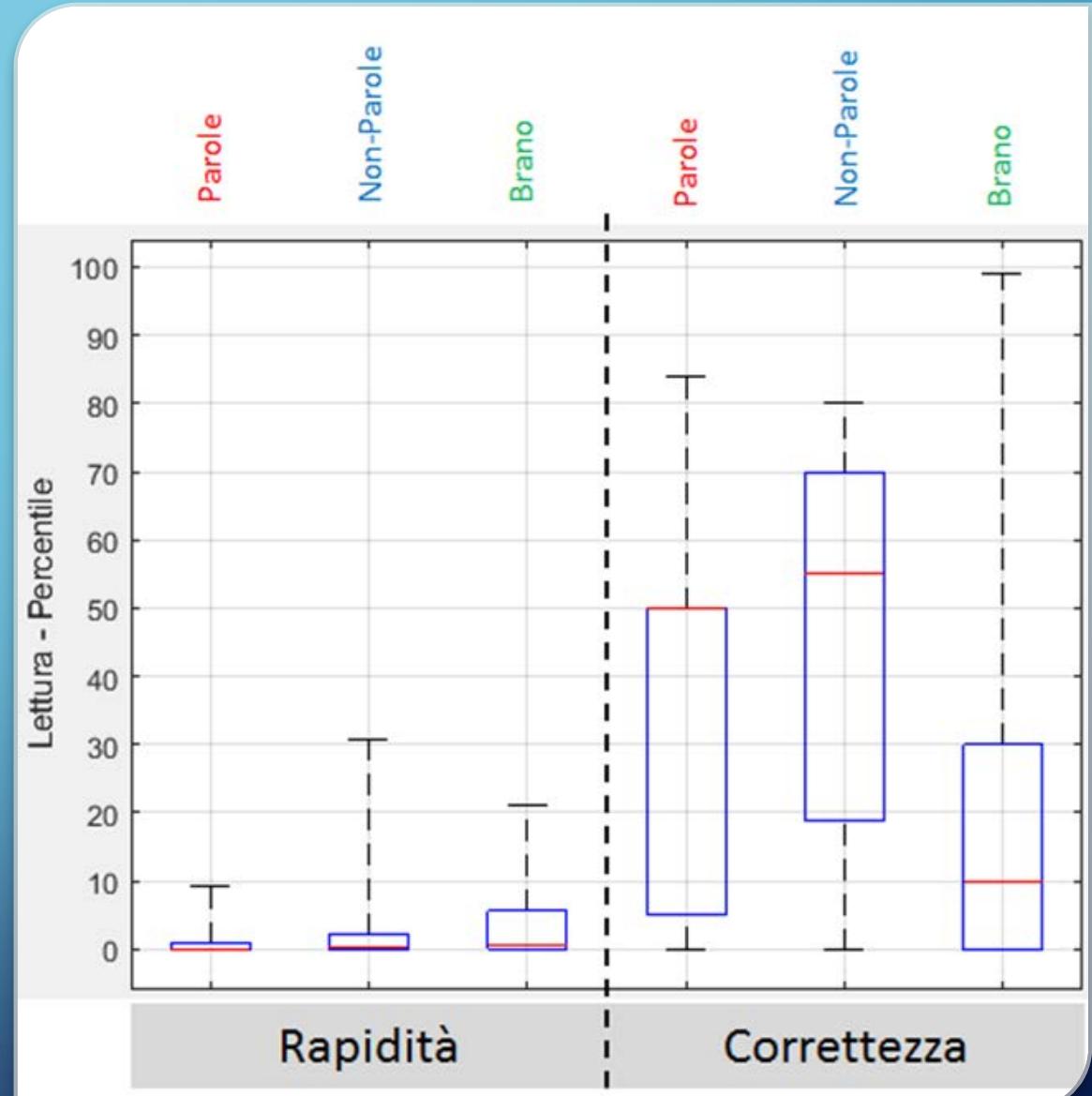
Questo vale, in particolar modo, per la lettura di non parole e di parole.



CARATTERISTICHE DELLA LETTURA NEI SOGGETTI CON DISLESSIA

Quindi, relativamente alla **rapidità di lettura**, i soggetti con dislessia risultano deficitari rispetto ai criteri di valutazione.

Per quanto riguarda, invece, la **correttezza**, le prestazioni dei soggetti con dislessia risultano deficitarie solo nel caso della lettura di brano.



CARATTERISTICHE DELLA SCRITTURA NEL CAMPIONE

Le prestazioni di scrittura, sia nel dettato di parole, sia nella scrittura di numeri in lettere **peggiorano notevolmente** se svolte in condizione di soppressione articolatoria.

La prova di scrittura di numeri in lettere diventa diagnostica del disturbo, per la maggior parte dei soggetti, solo nella condizione di soppressione articolatoria.

