

POLITECNICO DI MILANO
Corso di Studio in Ingegneria Biomedica



**PREDIZIONE DEL RISCHIO DI RITARDI
NELL'APPRENDIMENTO DELLA SCRITTURA DAL
PREGRAFISMO TRAMITE SERIOUS GAME SU TABLET**

Docente Corso Progetto: Simona Ferrante

Tutor: Simona Ferrante

Correlatore: Linda Greta Dui

Tesi di laurea di:

Peteani Giulia - Matr. 887483

Redaelli Niccolò - Matr. 888922

Rizzato Gloria - Matr. 887829

Ronchetti Francesca - Matr. 889833

Anno Accademico: 2019/2020

Indice

ABSTRACT	1
SOMMARIO	2
1. STATO DELL'ARTE	3
1.1 DISGRAFIA	3
1.2 CARATTERISTICHE DELLA SCRITTURA	4
1.2.1 DA CARTACEO A TAVOLETTA GRAFICA	4
1.2.2 CARATTERISTICHE DELLA SCRITTURA ATTRAVERSO L'UTILIZZO DI TECNOLOGIE	5
1.3 PRINCIPI CHE GOVERNANO L'ORGANIZZAZIONE RITMICA DELLA SCRITTURA	7
1.4 SPEED ACCURACY TRADE-OFF	7
1.5 OBIETTIVI	9
2. METODI E MATERIALI	10
2.1 PARTECIPANTI	11
2.2 PROTOCOLLO	11
2.3 APPLICAZIONE E STRUMENTI PER L'ACQUISIZIONE DATI	13
2.4 FEATURE ESTRATTE	14
2.5 ANALISI DEI FATTORI CONFONDENTI	17
2.6 FEATURE ENGINEERING	18
2.7 FEATURE SELECTION	19
2.8 TECNICHE ADDESTRAMENTO E VALUTAZIONE DEL MODELLO	20
3. RISULTATI	22
3.1 DESCRIZIONE DELLA POPOLAZIONE	22
3.2 ANALISI DEI FATTORI CONFONDENTI	24
3.3 IMPUTAZIONE E FEATURE ENGINEERING	26
3.4 CORRELAZIONE	28
3.5 ADDESTRAMENTO E VALUTAZIONE DEL MODELLO	30
4. DISCUSSIONE	33
4.1 ADDESTRAMENTO E VALUTAZIONE DEL MODELLO	33
4.2 ANALISI FATTORI CONFONDENTI	34
4.3 ANALISI FEATURE PIÙ PREDITTIVE	35
4.4 CONCLUSIONI	36
4.4.1 LIMITI DELL'ANALISI	36
4.4.2 SVILUPPI FUTURI	36
4.4.3 CONCLUSIONE FINALE	37
BIBLIOGRAFIA	38

Ringraziamenti

Vorremmo rivolgere un sentito ringraziamento al nostro tutor per averci dato l'opportunità di partecipare a questo progetto, e alla nostra correlatrice per essere stata sempre disponibile a condividere con noi le sue conoscenze e la sua esperienza personale, per averci guidato nel risolvere i problemi riscontrati durante il percorso spronandoci sempre a dare un'interpretazione ai risultati trovati con senso critico. Un grande ringraziamento va anche alle nostre famiglie e a tutti i nostri cari che ci hanno supportato in questo percorso.

ABSTRACT

Dysgraphia is a specific learning disability that affects the fundamental functions of handwriting in absence of any intellectual and neurological deficits. It is associated with difficulties in fine motor skills, often related to problems in praxic skills, which prevents to automate the motor routine necessary for the graphic gesture realization. By nowadays it is believed that a correct diagnosis of dysgraphia cannot be performed before the third grade.

Our study focuses on the possibility of anticipate the evaluation of the risk of a delay in learning handwriting in subject attending the last year of kindergarten, through a predictive model based on the analysis of pre-graphic exercises. Given that alterations of the handwriting rhythm and kinematic characteristics are observed in dysgraphic subjects, we have evaluated whether such alterations are confirmed in pre-graphic symbols, as well.

We leveraged data collected with serious games for iPad, aimed at testing different characteristics of handwriting through pregraphic symbols. Executions were compared with the teachers' judgement on children's graphical abilities. We focused on the analysis of alterations in the characteristics of handwriting and on the violations of the Isocrony and the Speed Accuracy Trade-off principles: features such as speed, performance's indices and frequency contents were extracted from data. The analysis of the confounders (age, gender, dominant hand, language, use of tablet) was carried out on these features, and datasets were created to train and evaluate a predictive model. Features engineering was leveraged to make features comparable by normalization. Feature selection was performed to selected those that would not introduce redundancy of information and correlated with the "risk" variable.

The impact of confounders was negligible, while a total of 22 features were selected as useful for the prediction. The predictive model was able to identify subjects at risk of delay in learning of handwriting, with a improvement of 25% compared to a baseline predictor.

The most relevant innovation of the work consists in the possibility to objectively predict children at risk of a delay in learning graphical abilities, excluding the subjective component introduced by the teachers' judgement, thanks to a quantitative analysis of pregraphic symbols, since pre-scholar age. In this way, those who will be identified as subjects at risk will have the opportunituy to undergo to an empowerment program and simultaneously will be monitored in order to observe their graphic abilities' evolution, even at distance of years, with the aim of confirm the predictions.

One limit of our analysis is represented by the restricted amount of available data, due to the historical moment we are experiencing, but we believe that it can be a great starting point for future deeper analysis.

SOMMARIO

La disgrafia è un disturbo specifico dell'apprendimento che incide sulle funzioni fondamentali della scrittura in assenza di deficit intellettivi e neurologici. È legata a difficoltà nella motricità fine, spesso associate a difficoltà nelle competenze prassiche, che impediscono di automatizzare la routine motoria necessaria per la realizzazione del gesto grafico. Ad oggi si ritiene che una diagnosi corretta della disgrafia non possa essere effettuata prima della terza primaria.

Il nostro studio si concentra sulla possibilità di anticipare la valutazione del rischio di ritardi nell'apprendimento della scrittura all'ultimo anno della Scuola dell'Infanzia, tramite un modello predittivo basato sull'analisi di esercizi pregrafici.

Sapendo che nei soggetti disgrafici si osservano alterazioni delle caratteristiche ritmiche e cinematiche della scrittura, nel nostro studio abbiamo valutato se queste vengono confermate anche nel caso dei simboli pregrafici.

I dati utilizzati sono stati raccolti tramite serious games per iPad, volti a testare diverse caratteristiche della scrittura tramite simboli pregrafici. Le esecuzioni sono state confrontate con il giudizio degli insegnanti sulle abilità grafiche dei bambini. Abbiamo posto maggiore attenzione sull'analisi di eventuali alterazioni delle caratteristiche della scrittura e sulle violazioni dell'Isocronia e della Speed-Accuracy Trade-off: sono state estratte feature quali, ad esempio, velocità, indici di performance e contenuti in frequenza a partire dai dati a nostra disposizione. Su tali feature è stata svolta un'analisi dei fattori confondenti (età, genere, mano dominante, lingua, uso tablet) e sono stati creati i dataset per l'addestramento e la valutazione di un modello predittivo. Le feature sono state rese comparabili attraverso la feature engineering (normalizzazione) e sono state selezionate in modo da utilizzare solo quelle correlate con la variabile "rischio" e che non introducessero ridondanza di informazioni.

L'impatto dei fattori confondenti è stato trascurabile, mentre un totale di 22 feature si è rivelato utile alla predizione. Si è poi ottenuto un modello predittivo in grado di identificare i soggetti a rischio di ritardi nell'apprendimento della scrittura con un miglioramento del 25% rispetto ad un predittore casuale.

La novità più rilevante dello studio consta nella possibilità di predire quali bambini siano a rischio di ritardi nello sviluppo delle abilità grafiche in modo oggettivo, escludendo quella componente di soggettività data dal giudizio degli insegnanti, sin da un'età prescolare, grazie all'analisi dei simboli pregrafici. In questo modo, coloro che verranno individuati come soggetti a rischio potranno essere sottoposti ad un percorso di potenziamento e parallelamente monitorati al fine di osservare l'evoluzione delle loro abilità grafiche, anche a distanza di anni, in modo da confermare le predizioni. Un limite della nostra analisi è da individuarsi nella quantità di dati a disposizione, fortemente limitata dal momento storico corrente, ma riteniamo che possa essere un ottimo punto di partenza per analisi future maggiormente approfondite.

1. STATO DELL'ARTE

1.1 DISGRAFIA

La disgrafia è un disturbo che comporta difficoltà nell'espressione scritta, cioè irregolarità per quanto riguarda la forma e la grandezza delle lettere, la spaziatura tra le lettere e tra le parole, lo spelling, la grammatica, la punteggiatura, l'impugnatura della penna, la velocità di scrittura e l'organizzazione spaziale [American Psychiatric Association, 2013]. La disgrafia fa parte dei Disturbi Specifici dell'Apprendimento (DSA) e in quanto tale, per definizione, non viene attribuita a soggetti che presentano ritardo mentale, problemi visivi e uditivi o disturbi psicologici ed emotivi.

A sua volta la disgrafia è articolata in tre varianti (dislessica, spaziale e motoria) basate sulle variabili *leggibilità della scrittura spontanea*, *leggibilità della scrittura da copiatura* e *velocità di scrittura*. Nella disgrafia dislessica viene compromessa solo la leggibilità della scrittura spontanea, nella disgrafia spaziale sia la scrittura spontanea che la copiatura sono illeggibili, ed infine nella disgrafia motoria oltre all'illeggibilità si osserva anche l'alterazione della velocità di scrittura [Deuel et al, 1995]. Particolare risulta essere il caso della diagnosi della disgrafia dislessica poiché la maggior parte dei test prevede la sola copiatura di parole, il che non mette in luce la presenza del disturbo.

La disgrafia evolutiva, in particolare, si sofferma sul concetto dell'ottenimento di scarsi risultati accademici seppur avendo potenzialmente buona capacità di apprendimento ed è stimata colpire tra il 5% e il 33% dei bambini in età scolare [Hamstra-Bletz et al, 1993; Karlsdottir et al, 2002; Smits-Engelsman et al, 2001]. Dati gli scarsi risultati scolastici, il bambino potrà subire ripercussioni per quanto riguarda il futuro accademico-lavorativo e potrebbe anche sviluppare problemi di autostima [Chung et al, 2020; Lyon et al, 1996; Feder et al, 2007]. Per evitare un effetto a cascata è quindi

importante diagnosticare il disturbo quanto prima in modo tale da applicare agevolazioni che diminuiscano lo svantaggio ed effettuare un potenziamento mirato per superare, dove risulta possibile, le difficoltà.

Una diagnosi corretta non può essere ottenuta prima della fine del secondo anno di scuola primaria – anche se alcuni studi affermano che sia meglio alla fine del terzo [Overvelde et al, 2011] –, quando i bambini raggiungono la completa maturazione della scrittura. In questo modo nei primi tre anni di istruzione elementare non vengono forniti gli strumenti appropriati per un apprendimento più efficace e recuperare comporta un grande lavoro da parte del bambino, degli insegnanti e della famiglia. Un’ottima strategia sarebbe quindi quella di limitare il tempo trascorso senza diagnosi lavorando in prevenzione: individuare i bambini a rischio di ritardi nell’apprendimento delle abilità grafiche attraverso l’analisi del gesto grafico relativo al disegno - inteso come precursore della scrittura - e, successivamente, effettuare un potenziamento con il fine di aiutarli a sviluppare quelle funzioni che stanno emergendo più lentamente rispetto ai coetanei.

1.2 CARATTERISTICHE DELLA SCRITTURA

Al giorno d’oggi la diagnosi viene effettuata con test soggettivi e poco precoci il che può portare incertezza sull’esito. L’attuale gold standard che permette di valutare la presenza o meno di disgrafia in modo qualitativo è il BHK Test [1987, Olanda]. In tale test al soggetto viene chiesto di copiare un testo su carta bianca in corsivo e successivamente vengono valutati diversi parametri che consentono di assegnare un punteggio per discriminare i soggetti disgrafici da quelli non disgrafici (tra questi vi sono parametri legati all’occupazione spaziale, allo spazio tra le parole, all’altezza delle lettere e al numero di auto-correzioni che vengono effettuate). Ideata sulla base della lingua olandese, questa scala è stata adeguata in modo da diventare fruibile anche in italiano.

1.2.1 DA CARTACEO A TAVOLETTA GRAFICA

Attualmente i test presenti per valutare la disgrafia sono cartacei, il che porta ad avere delle valutazioni non del tutto oggettive (basti pensare al punteggio finale assegnato, valutazione di uno specialista che seppur tenti di essere oggettivo, introdurrà sempre una variabilità intra e interoperatore). Da qui nasce dunque la necessità di utilizzare dei dispositivi che siano il più imparziali possibile, come ad esempio le tavolette grafiche [Danna et al, 2013; Mekyska et al, 2017] o tablet [Lunardini et al, 2015].

Le tavolette grafiche sono box digitalizzati in cui è possibile scrivere utilizzando una penna collegata ad esse. La limitazione di tale tecnologia consiste nel fatto che non si può vedere ciò che si sta scrivendo, infatti occorre un laptop per poter visualizzare il tratto.

Per ovviare a ciò, sulla superficie della tavoletta viene fissato un foglio di carta dove i partecipanti allo studio scrivono con una penna ad inchiostro, soluzione che però potrebbe portare ad una alterazione nelle misure (come ad esempio nella variabile pressione). Se si considerano invece i tablet, vengono utilizzate delle penne in grado esse stesse di misurare la pressione che poi verrà inviata al tablet e ciò permette di non utilizzare dei fogli di carta, poiché è possibile visualizzare quello che si sta scrivendo sulla superficie stessa del tablet.

1.2.2 CARATTERISTICHE DELLA SCRITTURA ATTRAVERSO L'UTILIZZO DI TECNOLOGIE

Con l'utilizzo di tali dispositivi, oltre a raggiungere una maggiore oggettività, è possibile distinguere i diversi tipi di disgrafia [Deuel et al, 1995] illustrati al paragrafo 1.1, infatti attraverso i test cartacei non è possibile riconoscere la disgrafia dislessica: il gold standard prevede solo la copiatura di un testo e non permette di valutare quantitativamente una scritta spontanea, condizione in cui si presenta tale declinazione del disturbo.

Vi sono altri fattori positivi nell'utilizzo di tavolette grafiche o tablet: oltre ad esaminare le caratteristiche proprie del prodotto finale sulle quali si basano i test cartacei, come ad esempio l'altezza delle lettere e la spaziatura tra due parole consecutive, è possibile valutare il processo della scrittura, ovvero prendere in considerazione la cinematica e la dinamica del gesto grafico.

A tale proposito molti sono gli studi che hanno come obiettivo lo sviluppo di un modello predittivo per la valutazione delle abilità grafiche utilizzando le tecnologie sopra riportate. Per raggiungere questo scopo, si ricorre all'estrazione di diversi parametri della scrittura che possono essere raggruppati in tre macrocategorie: caratteristiche del prodotto finale, caratteristiche variabili nel tempo e caratteristiche relative al contenuto in frequenza.

Per quanto riguarda la prima categoria - caratteristiche del prodotto finale - vengono prese in considerazione caratteristiche legate all'allineamento, all'altezza e allo spazio fra le parole [T. H. Falk et al, 2011].

Dalla definizione di disgrafia si evince che i soggetti disgrafici tendono ad avere difficoltà nel rispettare l'altezza delle lettere.

Per la seconda macrocategoria - caratteristiche variabili nel tempo - le più rilevanti risultano essere quelle legate alla velocità di scrittura, come ad esempio velocità media, massima e la sua deviazione standard, ma anche quelle legate alla velocità con cui la penna cambia accelerazione (Jerk). Risultano avere importanza anche le caratteristiche legate alla pressione, quali pressione media, massima e la sua deviazione standard. Altre caratteristiche rilevanti sono: In Air Time che rappresenta la porzione di tempo trascorsa senza toccare la superficie del tablet/tavolettina grafica e numero di interruzioni, ovvero quante volte il soggetto stacca la penna dalla superficie. Nei soggetti disgrafici le velocità medie sono più basse e le velocità massime più alte [T. Asselborn, et al, 2020]; la pressione ha valori più elevati [J. Mekyska et al, 2017] e In Air-Time maggiore [T. Asselborn, et al, 2018; Rosenblum et al, 2003].

Per le caratteristiche legate al contenuto in frequenza le più rilevanti nei diversi studi sono: la mediana della potenza spettrale delle frequenze di velocità, di inclinazione della penna rispetto alla direzione di scrittura (tilt-x), di inclinazione della penna perpendicolarmente alla direzione di scrittura (tilt-y), di velocità con cui varia la pressione e Signal to Noise velocity peaks difference (SNvpd) [Danna et al, 2013]. Quest'ultima rappresenta la differenza tra il numero di picchi di velocità una volta filtrata con una frequenza di taglio di 5 Hz e una volta con frequenza di taglio di 10 Hz (banda tra 0 e 5 Hz rappresentante il movimento controllato e banda tra 5 e 10 Hz rappresentante le fluttuazioni non controllate presenti nel sistema motorio) [Teulings et al, 1984].

Nei soggetti disgrafici risulta che le frequenze di velocità sono più elevate: sono presenti infatti cambiamenti rapidi di velocità, accelerazioni e decelerazioni, tradotti in alte frequenze. Ciò risulta essere correlato alle saccadi, ovvero ai rapidi movimenti oculari [Asselborn et al, 2018]. Tilt-y si mantiene costante il che riflette la rigidità con la quale il soggetto disgrafico impugna la penna, e tilt-x non risulta essere costante, infatti un soggetto non disgrafico ha più controllo nella direzione di scrittura. Il numero di fluttuazioni di velocità nelle bande 0-5 Hz e 5-10 Hz non si mantiene costante, il che rappresenta maggior rumore nel sistema neuromotorio, manifestando così irregolarità nel controllo del movimento.

Sebbene l'obiettivo di molti studi risulti essere lo stesso, ovvero quello di discriminare i soggetti disgrafici dai non disgrafici, spesso ci sono delle discrepanze: in alcuni studi infatti risulta che le caratteristiche legate alla velocità di scrittura siano quelle che riescono meglio a separare le due categorie [Asselborn et al, 2018] e in altri quelle legate all'inclinazione della penna [J. Mekyska et al, 2017], in altri ancora quelle inerenti alla pressione [Asselborn et al, 2020].

1.3 PRINCIPI CHE GOVERNANO L'ORGANIZZAZIONE RITMICA DELLA SCRITTURA

La scrittura a mano non è semplicemente una successione di atti motori isolati, bensì è un processo organizzato, gerarchico e ritmico, in cui il tempo e lo spazio di ciascuna unità motoria (tratti, lettere) sono interdipendenti tra loro nell'unità più grande (parola) [Lashley, 1951; Fitch et al, 2014; Van Galen et al, 1983]. Tra i principi che governano la scrittura, regolandone l'organizzazione temporale dei singoli eventi motori, vi è l'Isocronia.

Con Isocronia si intende la relazione proporzionale tra la velocità di esecuzione e la lunghezza della traiettoria del tratto. Significa che il tempo impiegato nello scrivere una parola resta costante al variare delle condizioni di scrittura imposte [Binet et al, 1893; Stetson et al, 1923; Viviani et al, 1982]. La tendenza è quella di scrivere più velocemente dovendo percorrere un tratto più ampio e viceversa. Ciò viene fatto al fine di mantenere costante la durata dell'esecuzione.

Generalmente violazioni di Isocronia sono riscontrate nei soggetti disgrafici. Ad esempio, nonostante considerevoli aumenti o riduzioni dell'ampiezza del tratto, se non dovessero verificarsi notevoli variazioni della velocità di scrittura, tale principio verrebbe violato [Pagliarini et al, 2015].

1.4 SPEED ACCURACY TRADE-OFF

Nella valutazione della performance di un atto motorio è necessario considerare due parametri: la velocità e l'accuratezza dell'esecuzione, indicativi della capacità di controllo motorio di un soggetto.

Il sistema psicomotorio è un sistema intrinsecamente rumoroso: l'esecuzione di un comando avviene attraverso le informazioni sensoriali elaborate dai componenti del sistema motorio che, a causa della sua natura, ha una capacità limitata di trasporto dei segnali. Tale limite presente nella propagazione delle informazioni è alla base del compromesso tra accuratezza e velocità di un atto motorio [Van Galen et al, 1995].

Questo compromesso accomuna molte esecuzioni motorie: più l'atto viene eseguito con precisione, più tempo richiede la sua esecuzione.

La prima legge formulata per descrivere questo compromesso è la legge di Fitts nel 1954 che predice la performance di movimenti mirati, volti ad un bersaglio. La legge è rappresentata dalla seguente espressione:

$$MT = a + b \cdot \log_2 \frac{2 \cdot A}{W} \quad (1)$$

MT è il tempo impiegato per raggiungere un obiettivo di larghezza W a distanza A; a e b sono due costanti empiriche dipendenti dall'individuo che esegue il movimento.

La legge di Fitts permette di calcolare l'indice di difficoltà ID dell'esecuzione in funzione dei parametri geometrici A e W:

$$ID = \log_2 \frac{A}{W} \quad (2)$$

Nel nostro caso, però, vengono posti sotto studio atti motori che rimandano al gesto grafico della scrittura o del disegno, consistenti nel tracciamento di traiettorie che, quindi, non possono essere modellati da tale legge. Nel 1997 Accot e Zahi condussero uno studio per trovare l'esistenza di regolarità in compiti basati sul tracciamento di percorsi, con risultato finale la Steering Law. La legge permette di calcolare il tempo impiegato per tracciare un percorso rimanendo attentamente entro i bordi di un tunnel di dimensioni definite; si ottiene dalla Legge di Fitts, applicando lungo il tunnel un numero di bersagli tendente all'infinito, come mostrato in Figura 1 [Accot, 1997].

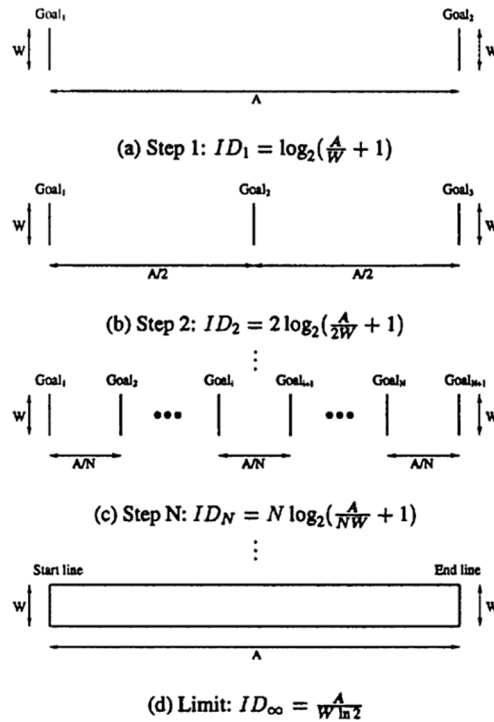


Figura 1: la Figura mostra i passaggi che permettono di ottenere la Steering Law a partire dalla Legge di Fitts.

In questo modo si ottiene una nuova espressione dell'indice di difficoltà ID:

$$ID = \frac{A}{W \cdot \ln 2} \quad (3)$$

Che per semplicità viene approssimata nel seguente modo:

$$ID = \frac{A}{W} \quad (4)$$

È intuibile quindi che la relazione che lega la durata dell'esecuzione con l'indice di difficoltà è di tipo lineare, non più logaritmica, ed è data dalla seguente espressione:

$$MT = a + b \cdot \frac{A}{W} \quad (5)$$

La Steering Law permette di stimare il tempo di esecuzione MT, a partire dai parametri A e W che, in questo caso, rappresentano rispettivamente la lunghezza e la larghezza del tunnel nel quale viene eseguita la traiettoria.

MT è una variabile dipendente dall'indice di difficoltà ID imposto, dato dal rapporto tra la lunghezza e la larghezza del tunnel: più grande è l'indice di difficoltà, maggiore risulta essere il grado di difficoltà del compito.

1.5 OBIETTIVI

Dalla letteratura emerge la necessità di una diagnosi precoce della disgrafia in età prescolare, al fine di intervenire preventivamente: l'obiettivo di questo progetto è la creazione di un modello predittivo capace di individuare i bambini a rischio di ritardi nell'apprendimento delle abilità grafiche, attraverso l'analisi delle caratteristiche dell'esecuzione del gesto grafico. In questo modo tali soggetti potranno essere indirizzati e guidati verso un percorso di potenziamento e un monitoraggio dell'evoluzione delle loro abilità grafiche.

Sono stati prefissati anche due obiettivi secondari, ovvero l'analisi dei fattori confondenti e l'identificazione delle feature più predittive. Per il primo viene fatta una valutazione di tutti quei fattori quali genere, età, mano dominante, madrelingua e dimestichezza nell'utilizzo del tablet che avrebbero potuto influenzare i risultati del modello. Per il secondo invece, sono state estratte dai dati

feature rappresentative dell'atto motorio della scrittura, su cui è stata eseguita un'operazione di selezione al fine di individuare quelle che risultano essere più predittive.

2. METODI E MATERIALI

Per raggiungere gli obiettivi prefissati abbiamo seguito l'iter riportato in Figura 2 tratto da Mekyska et al, 2017 e riadattato per questo progetto. Partendo da dati forniti, il primo step è stato l'estrazione delle feature – elencate al paragrafo 2.4 – per i tre esercizi presi in considerazione, ovvero copia quadrato, tunnel quadrato e tunnel “ele”. Su queste feature estratte è stata eseguita la feature engineering (paragrafo 2.6) al fine di ridimensionarle e renderle comprensibili alla macchina. Il passo successivo è stato l'analisi dei fattori confondenti utilizzando i test esposti al paragrafo 2.5, da cui sono state estratte conclusioni sulla loro influenza e successivamente confrontate con quanto appreso dalla letteratura. Si è passati poi alla feature selection (paragrafo 2.7) in cui è stato ridotto il numero delle feature valutandone la correlazione, svolta al fine di stabilire eventuali relazioni lineari tra feature e la variabile rischio e successivamente tra le feature stesse. Lo step finale è stato l'addestramento e la valutazione di vari modelli tra i quali è stato scelto il migliore (paragrafo 2.8), basandosi principalmente sulla metrica F1-Score. A questo punto si è ottenuto un modello in grado di effettuare una classificazione finale identificando i soggetti che dovranno essere sottoposti a potenziamento e, parallelamente, monitorati al fine di osservare l'evoluzione delle loro abilità grafiche a distanza di anni.

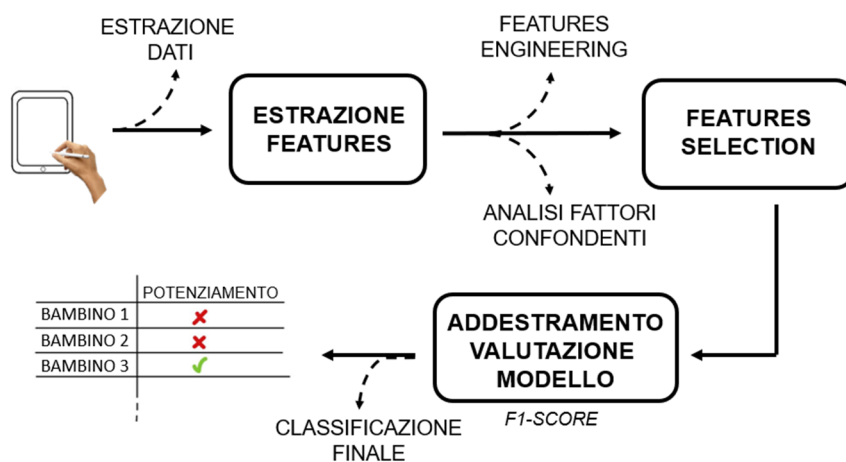


Figura 2: schema degli step effettuati per la creazione del modello predittivo approfonditi nei paragrafi successivi.

2.1 PARTECIPANTI

Per lo studio sono stati reclutati 247 bambini frequentanti l'ultimo anno della Scuola dell'Infanzia in istituti della provincia di Como e di Varese prendendo come criterio di esclusione la diagnosi di patologie neurologiche, visive o motorie che possano compromettere la scrittura. A tutti i soggetti risultati idonei sono state richieste delle informazioni personali (previo il consenso dei genitori al trattamento di tali dati) ed è stata assegnata una classificazione da parte delle insegnanti relativamente alla presenza del rischio di ritardi nello sviluppo delle abilità grafiche.

Il range di incidenza della disgrafia previsto è tra il 5 e il 33% [Overvelde, 2011, Reynolds, 2007], Figura 3. È stato considerato il caso peggiore - ovvero 5% - e perciò per trovare almeno 10 soggetti a rischio e garantire un dropout del 20%, cioè con la possibilità che 50 si possano ritirare, sono stati reclutati circa 250 bambini.

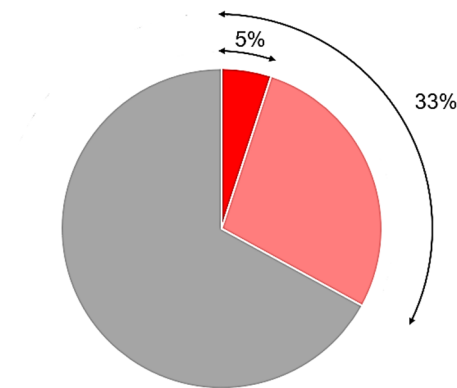


Figura 3: range incidenza prevista, in rosso e rosa i soggetti a rischio, in grigio quelli non a rischio.

2.2 PROTOCOLLO

La raccolta dati prevede diverse fasi in cui ai soggetti viene richiesto di svolgere varie attività, quali la copiatura di simboli pregrafici e il tracciamento di traiettorie all'interno di tunnel di diverse forme. Di seguito vengono riportate in modo più dettagliato le prove effettuate dai soggetti.

- Copiatura di una figura pregrafica (quadrato) in tre condizioni differenti: spontaneo (esempio riportato in Figura 4), grande e piccolo. La finalità di questo esercizio è di valutare il principio di Isocronia tra le diverse modalità;

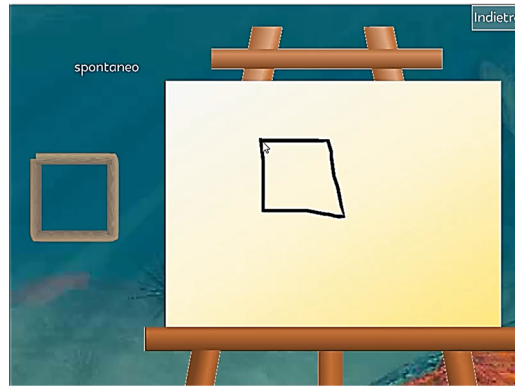
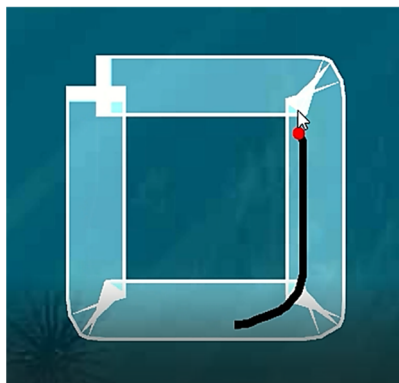
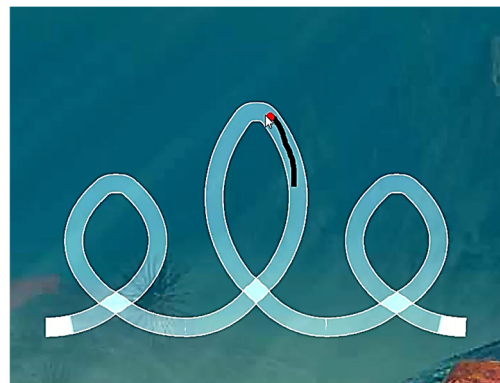


Figura 4: esercizio copia quadrato – condizione spontanea.

- Prova di disegno all'interno di due diversi percorsi prestabiliti: tunnel quadrato in Figura 5a e tunnel “ele” in Figura 5b. Il primo corrisponde ad un tunnel di forma quadrata, mentre il secondo ad un tunnel che ricorda la sequenza delle lettere e-l-e. Questa prova consiste nel tracciamento di traiettorie all'interno di 14 tunnel di lunghezza e larghezza variabili per ciascuna tipologia di percorso, per ottenere 5 diverse combinazioni di difficoltà ID date dal rapporto tra la lunghezza (A) e la larghezza (W). Queste sono state eseguite al fine di valutare la Speed Accuracy Trade-off.



(a)



(b)

Figura 5a-5b: esercizio tunnel quadrato a sinistra, ed esercizio tunnel “ele” a destra.

Parallelamente sono state oggetto di analisi anche alcune caratteristiche della scrittura presenti in letteratura per quanto riguarda il prodotto finale, oltre che alle caratteristiche variabili nel tempo e quelle legate al contenuto in frequenza. Al fine di poter effettuare tali valutazioni, per tutte le prove sono stati raccolti dati di posizione, pressione e inclinazione della penna, associati alla frequenza di campionamento di 250Hz.

2.3 APPLICAZIONE E STRUMENTI PER L'ACQUISIZIONE DATI

I dati a noi forniti sono stati raccolti tramite l'applicazione *Play Draw Write*, realizzata col motore grafico multiplatforma Unity che consente lo sviluppo di videogiochi e contenuti interattivi. L'interfaccia grafica propria del game design ha reso l'attività ai soggetti più “leggera e familiare”.

Tale applicazione è stata installata su degli iPad dati in uso ai bambini, attraverso cui hanno potuto svolgere gli esercizi proposti. Gli iPad in dotazione erano di VI o VII generazione e le differenze tecniche che hanno influito maggiormente nella nostra analisi sono le seguenti:

- Il dispositivo di VI generazione dispone di un display Multi-Touch retroilluminato LED da 9,7" con tecnologia IPS, mentre quello di VII generazione di un display della stessa tipologia ma da 10,2";
- L'iPad di VI generazione ha una risoluzione di 2048x1536 pixel a 264 ppi (pixel per pollice), mentre quello di VII ha una risoluzione superiore pari a 2160x1620 pixel a 264 ppi.

Entrambi i tablet supportano l'Apple Pencil di prima generazione che in alcuni casi potrebbe addirittura essere considerata un fattore incisivo sulle “prestazioni” dei bambini. Tale assunzione è fatta per via delle sue dimensioni e del relativo peso, rispettivamente 17,57 cm e 20,7 g, che potrebbero aver reso complicata l'impugnatura ai soggetti. La penna contiene poi un accelerometro Bosh Sensortech BMA280 3-Axis e sensori in grado di rilevare pressione e angolatura della penna sulla superficie del tablet. Tutti i dati che abbiamo utilizzato sono stati campionati a 250Hz.

2.4 FEATURE ESTRATTE

Per rendere il nostro studio il più completo possibile sono state estratte 138 feature, in particolar modo 78 feature per l'esercizio copia quadrato nelle tre diverse condizioni (spontanea, grande, piccola), mentre le restanti 60 equamente suddivise tra l'esercizio tunnel quadrato e tunnel "ele".

Iniziando dall'esercizio copia quadrato, in Tabella 1 sono riportate le feature estratte.

Categoria feature estratte	...per valutare
Velocità	Isocronia
Fluidità	Caratteristiche della scrittura
Padronanza	
Occupazione spaziale	
Frequenza	

Tabella 1: caratteristiche estratte e rispettivo scopo di valutazione per esercizio copia quadrato.

Al fine di valutare il principio di Isocronia sono state estratte feature riguardanti la velocità, quali:

- *Velocità media e velocità massima* nelle tre diverse condizioni di esercizio (spontaneo, grande, piccolo) al fine di verificare se la velocità nelle tre modalità variasse in accordo col principio di Isocronia. In particolare, sono state calcolate le velocità istantanee prendendo in considerazione il tempo totale di esecuzione di ciascuna prova, applicando un filtro a media mobile a 5 campioni. Per la velocità media è stato utilizzato un filtro passa basso con frequenza di taglio 10 Hz, mentre è stata definita velocità massima la media del 10% dei valori massimi assunti dalla velocità precedentemente estratta.

Passando poi all'analisi delle caratteristiche della scrittura, sono state estratte feature riguardanti la fluidità, la padronanza, l'occupazione spaziale e il contenuto in frequenza di diversi segnali relativi al gesto grafico nelle tre diverse condizioni di esercizio.

Per quanto riguarda la fluidità, le feature estratte sono:

- *Deviazione standard velocità*;
- *SNvpd* (Signal to Noise velocity peaks difference) in cui per prima cosa viene filtrata la velocità con un filtro Butterworth del quarto ordine e successivamente viene applicato un

filtro passa basso. Poi sono stati calcolati i picchi di velocità con le relative posizioni. Tali passaggi sono stati ripetuti due volte, considerando prima come frequenza di taglio 5 Hz e poi 10 Hz. In conclusione, è stata calcolata la differenza tra il numero di picchi di velocità ottenuti con le due diverse frequenze di taglio;

- *Tempo trascorso sotto una soglia di velocità*, dove in particolare la soglia è definita come la somma tra il valore minimo assunto dalla velocità e il 3% del range della velocità.

Per la padronanza le caratteristiche estratte sono:

- *In air-time* in cui sono stati sommati tutti gli intervalli di tempo in cui la penna non tocca la superficie del tablet;
- *Varianza azimuth e varianza altitude* dove per azimuth (Figura 6a) si intende l'inclinazione della penna rispetto alla direzione di scrittura e per altitude (Figura 6b) l'inclinazione della penna perpendicolarmente alla direzione di scrittura.

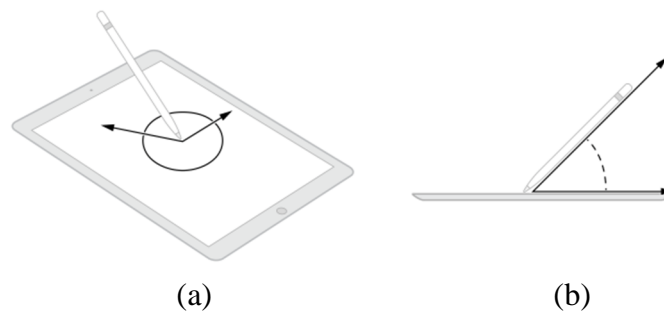


Figura 6a-6b: rappresentazione grafica di azimuth e altitude.

Per l'occupazione spaziale sono state considerate:

- *Centroide X e centroide Y*, ovvero le mediane relative alle coordinate x e y della traiettoria disegnata dal soggetto riferite alla superficie del tablet.

Infine, in riferimento al contenuto in frequenza per pressione, azimuth, altitude e velocità sono state estratte:

- *Ampiezza della banda* con soglia pari a 0.1;
- *Ampiezza e posizione primo picco*;
- *Energia del segnale* contenuta nella banda da 0 a 10 Hz.

Passando poi agli esercizi tunnel quadrato e tunnel “ele”, in Tabella 2 sono riportate le feature estratte.

Categoria feature estratte	Valutazione di
Tempo di percorrenza	Speed Accuracy Trade-off con Steering Law
Errore quadratico medio	
R^2	
Indice di performance	
Fluidità	Caratteristiche della scrittura
Padronanza	
Contenuto in frequenza	

Tabella 2: caratteristiche estratte e rispettivo scopo di valutazione per esercizi tunnel quadrato e tunnel “ele”.

Per gli esercizi consistenti nel tracciamento di percorsi all’interno di tunnel di forma quadrata e di forma che replica la sequenza di lettere e-l-e sono state estratte diverse caratteristiche per valutare la Speed Accuracy Trade-off con la Steering Law, quali:

- *Tempo di percorrenza* in cui prima è stato calcolato il tempo impiegato dal soggetto per tracciare ogni singolo tunnel e successivamente è stata considerata come caratteristica la mediana di questi tempi;
- *Errore quadratico medio* il quale indica l’errore commesso dalla regressione tra gli indici di difficoltà e il tempo impiegato dal soggetto per completare il tracciamento dei tunnel;
- R^2 , indice di bontà della regressione;
- *Indice di performance*, ovvero l’inverso del coefficiente angolare della retta di regressione;
- *Significant*, ovvero un booleano che rappresenta se la regressione è significativa o meno.

Per quanto riguarda la valutazione delle caratteristiche della scrittura sono state considerate le stesse feature utilizzate anche nell’esercizio copia quadrato, aggiungendone alcune relative alla fluidità. Risulta necessaria una distinzione tra i due tunnel; per il tunnel di forma quadrata sono state estratte:

- *Velocità media nei tratti interni al tunnel e velocità media nei tratti esterni al tunnel*;
- *Percentuale di traiettoria esterna al tunnel* calcolata come il rapporto tra la lunghezza della traiettoria esterna al tunnel e la lunghezza della traiettoria totale tracciata dal soggetto.

Mentre per il tunnel “ele” sono state considerate:

- *Numero errori* in cui sono stati calcolati il numero di incroci nella parola ele - tra le lettere e-l-e scritte in corsivo - non eseguiti correttamente dal soggetto nel tracciare la traiettoria;
- *Numero esecuzioni errate*, ovvero il numero di esecuzioni in cui il bambino ha commesso almeno un errore nel percorrere il tunnel.

Per quanto riguarda le caratteristiche estratte per i due tunnel è stata considerata solo la prova con lunghezza del percorso maggiore e ampiezza minore, ovvero la più difficoltosa, ad eccezione delle caratteristiche *numero errori* e *numero esecuzioni* che sono state valutate su tutte le prove.

Come è probabile che accada, alcune delle feature estratte possono presentare dei missing, ovvero per alcuni soggetti non è possibile calcolarne il valore. Per trattare questa mancanza di dati si procede in base alla percentuale di missing relativa alla feature: nel caso in cui sia minore del 5% la mediana dei valori presenti per quella feature viene inserita al posto dei valori mancanti; nel caso di feature con percentuale di missing maggiore del 5% si applica la regressione attraverso il toolbox di Matlab “Regression Learner”. Grazie ad esso è possibile addestrare vari tipi di modelli di regressione per poi scegliere quello avente il minor errore quadratico medio (RMSE), attraverso cui predire il valore mancante sulla base di quelli assunti dalle altre feature relative al soggetto con missing.

2.5 ANALISI DEI FATTORI CONFONDENTI

A partire dalle feature estratte viene condotta un’analisi relativa ai fattori confondenti. Lo scopo di tale analisi è quello di individuare se differenze nell’età dei bambini, assenza di lateralità definita per quanto riguarda la mano dominante, utilizzo pregresso del tablet, mancata comprensione della lingua possano influenzare le feature.

Per testare l’effetto delle suddette variabili su ciascuna feature si eseguono test statistici di ipotesi. Per prima cosa è necessario osservare la distribuzione dei valori della variabile dipendente: si esegue il test statistico Lilliefors Test, test di normalità per verificare l’ipotesi nulla che i dati provengono da una popolazione normalmente distribuita, di cui i parametri sono incogniti. In base alla distribuzione di dati si sceglie la statistica parametrica, che assume che i dati siano tratti da una popolazione normalmente distribuita, o la statistica non parametrica, che non fa alcuna assunzione riguardo la distribuzione della popolazione.

Nel caso in cui si testì l’effetto di una variabile indipendente continua (l’età) su una feature continua, qualora questa sia distribuita normalmente si utilizza la Correlazione di Pearson: il coefficiente di correlazione lineare di Pearson è un valore compreso tra -1 e 1. Se è pari a 0 non vi è una relazione

lineare tra le variabili, se pari a ± 1 le variabili sono perfettamente correlate rispettivamente positivamente (variano in modo concorde) e negativamente (variano in modo discorde). Viceversa, si usa la Correlazione di Spearman: il coefficiente di Spearman (ρ) è compreso tra -1 e 1 e si utilizza su dati in cui l'unica ipotesi fatta è quella che essi siano una misura ordinale.

Se invece si testa l'effetto di una variabile indipendente categorica a 2 livelli (il genere, la dimestichezza con il tablet e madrelingua) su una feature continua, nel caso in cui quest'ultima sia distribuita normalmente si usa il Test di Student (Test T) dividendo i dati in due classi a seconda della variabile di stratificazione. Il Test T verifica se i dati provengono da distribuzioni indipendenti con uguale media e varianza. In caso contrario si usa il Mann-Whitney Test, test non parametrico che verifica se i dati provengono da distribuzioni indipendenti continue con mediane uguali.

Tramite il p-value, il test indica se la variabile di stratificazione ha effetto sulla variabile che si testa. La soglia di significatività alpha è stata fissata a 0.05.

Infine, per testare l'effetto di una variabile categorica a più livelli (mano dominante) su una feature continua si utilizza in caso di normalità il One-Way-ANOVA altrimenti il Kruskal-Wallis Test. L'analisi della varianza (ANOVA) permette di confrontare più gruppi di dati comparando la variabilità interna a ciascun gruppo con la variabilità fra i gruppi: verifica che i dati di tutti i gruppi abbiano la stessa origine, ovvero provengano dalla stessa distribuzione stocastica. In caso di differenze fra gruppi, per sapere quali sono diversi tra loro, si usano i post-hoc-comparison, ovvero si confrontano tutte le coppie di gruppi con delle statistiche ad hoc (es. T-Test) con correzione di Bonferroni.

Oltre al test sulla singola variabile indipendente categorica, esclusivamente con distribuzioni normali, si può utilizzare l'Anova Multi-Via, test che permette di testare più variabili indipendenti insieme e capirne l'interazione.

2.6 FEATURE ENGINEERING

Coerentemente con l'obiettivo di creare un modello predittivo, l'operazione successiva da eseguire sulle feature estratte consiste nella feature engineering, svolta al fine di ridimensionarle per portarle sulla stessa scala.

Allo scopo di portare le feature continue ad un range simile per poterle comparare si utilizza la tecnica dello Scaling secondo cui è possibile applicare due procedure: la normalizzazione, tramite la quale i

dati vengono ridimensionati su un intervallo fisso, da 0 a 1, oppure la standardizzazione, metodo con cui si ottiene una distribuzione di dati con valore medio pari a 0 e deviazione standard pari a 1.

Per trattare le variabili categoriche a più livelli, invece, si esegue la tecnica del One Hot Encoding: si rimuove la variabile con codifica intera e viene aggiunta una nuova variabile binaria per ciascun valore intero biunivoco.

2.7 FEATURE SELECTION

Dopo aver svolto la feature engineering risulta necessario selezionare le feature. La selezione delle caratteristiche permette di identificare e rimuovere informazioni irrilevanti e ridondanti prima che avvenga la fase di apprendimento, altrimenti si potrebbe andare incontro al fenomeno di *overfitting* in cui il modello si adatta molto bene alle caratteristiche dei dati di addestramento, anche se queste potrebbero non avere alcun riscontro su dati mai esplorati. Tale modello perciò non è generalizzato, ovvero non riesce a predire la classe di appartenenza dei dati che non fanno parte dell'addestramento.

È possibile fare ulteriori osservazioni su tale fenomeno prendendo in considerazione rispettivamente la complessità del modello e l'errore commesso sui dati di addestramento e sui dati di validazione. All'aumentare della complessità del modello la percentuale di errore commessa sui dati forniti nella fase di addestramento diminuisce, mentre sui dati di validazione aumenta, dunque si verifica il fenomeno di *overfitting*. Se invece la complessità del modello diminuisce troppo, la percentuale di errore commesso sui dati sarà in entrambe le due fasi elevata: si va incontro al fenomeno opposto, ovvero l'*underfitting*. La situazione ideale a cui si aspira risulta essere un bilanciamento tra complessità del modello ed errore commesso: questo, oltre a migliorare le prestazioni del modello e permettere tempi più brevi e meno costosi a livello computazionale, è uno tra gli obiettivi della selezione delle caratteristiche.

A tal proposito esistono diversi metodi per eliminare le caratteristiche ridondanti, tra i quali la correlazione, sia tra le feature che tra la variabile da predire e le feature, e l'analisi delle componenti principali (PCA). In particolare, si parla di *feature selection* quando si eliminano le feature, mentre di *dimensionality reduction* nel momento in cui si riduce la complessità del modello senza eliminare le feature.

Per quanto riguarda la correlazione, essa permette di stabilire la relazione lineare che intercorre tra due variabili. In particolare, se la correlazione è significativa (nel caso di $p\text{-value} \leq 0.05$) e bassa (coefficiente di correlazione minore di 0.1) le caratteristiche sono indipendenti, al contrario se è significativa ed elevata (coefficiente di correlazione maggiore di 0.9) le due variabili non aggiungono

maggiori informazioni se considerate entrambe, infatti esse non sono indipendenti ed è perciò possibile predirne una dall'altra. In conclusione, vengono selezionate le caratteristiche che hanno una correlazione minore di 0.9, ovvero quelle che correlano poco tra di loro. Se si pone invece l'attenzione sulla correlazione tra la variabile da predire e la caratteristica presa in esame, questa non deve essere eccessivamente bassa (maggiore di 0.1) poiché il modello deve addestrarsi per poter classificare nel modo giusto i dati.

L'analisi delle componenti principali si occupa di ridurre la dimensionalità dello spazio predittore combinando linearmente le caratteristiche iniziali. La trasformazione lineare che si applica alle feature proietta le nuove variabili su diversi assi cartesiani: sul primo asse la variabile prima per varianza, sul secondo asse la variabile seconda per varianza e così per le rimanenti. Mediante tale procedura è possibile calcolare per ogni componente principale la percentuale di varianza spiegata rispetto alla varianza complessiva. Come già descritto, la prima componente principale è quella che spiega più varianza, seguono poi le altre in ordine decrescente. Se invece si considera la curva che esprime la varianza cumulativa, è possibile notare che questa assume la forma a gomito oltre il quale la varianza cumulativa incrementa di un valore non significativo.

Per il nostro progetto dal punto di vista computazionale l'analisi è stata eseguita attraverso MATLAB®. In particolare, all'interno del Classification Learner (toolbox di Matlab) è stata abilitata la PCA.

2.8 TECNICHE ADDESTRAMENTO E VALUTAZIONE DEL MODELLO

Per la creazione del modello predittivo ci si è basati su sistemi di machine learning che cercano di trovare relazioni tra le caratteristiche e le classi fornite per creare un modello durante la fase di addestramento al fine di poterlo utilizzare per classificare nuovi dati [Hall et al, 1998]. Si è partiti dal dataset a nostra disposizione, consistente nell'insieme di osservazioni ognuna con i relativi input (valori delle feature estratte) e il relativo output (classe di rischio fornita dalle insegnanti), applicando i seguenti step:

- Ordinamento casuale delle osservazioni appartenenti al dataset: per evitare di ottenere set contenenti tutti i bambini frequentanti la stessa scuola;
- Divisione del dataset: 2/3 destinati al training-set per l'addestramento del modello e il restante 1/3 utilizzato come test-set per la valutazione finale del modello completamente addestrato;
- Addestramento del modello predittivo: individuazione del valore dei parametri che definiscono il modello a partire dalla ricerca di nessi tra input e output. Per questo step è stato utilizzato il toolbox di Matlab "Classification Learner", un'applicazione che addestra

modelli di classificazione di dati tramite algoritmi di machine learning; consente di esplorare i dati, selezionare funzionalità, specificare schemi di validazione, formare modelli e valutare i risultati. Tramite l'applicazione si può eseguire un addestramento automatizzato per cercare il miglior tipo di modello di classificazione, scegliendo tra *decision tree*, *discriminant analysis*, *support vector machines*, *logistic regression*, *nearest neighbors*, *naive Bayes*, ed *ensemble classification*. Per utilizzare il modello con nuovi dati è possibile esportarlo nell'area di lavoro o generare il codice MATLAB® per ricreare il modello addestrato. Inoltre, per visualizzare i risultati del processo, il toolbox di Matlab contiene strumenti quali lo scatter plot, la matrice di confusione e la percentuale di accuratezza relativi.

All'interno del Classification Learner è stata abilitata la tecnica *10-fold-crossvalidation*; la crossvalidazione consiste nel suddividere il training-set in k sottogruppi (in questo caso 10) tutti con lo stesso numero di osservazioni e a turno k – 1 sottogruppi vengono usati per addestrare il modello, mentre il sottogruppo rimanente viene usato come set per la validazione - ottimizzazione dei parametri del modello ricavati attraverso l'addestramento -. Questa operazione di addestramento-validazione quindi si ripete k volte in modo tale che ogni sottogruppo venga usato esattamente una volta come validation-set. I k risultati vengono poi mediati per produrre una singola stima del valore dei parametri che definiscono il modello;

- Valutazione: sottoporre al modello predittivo degli input mai esplorati per poi giudicare la validità della classificazione fatta, paragonandola agli output corrispondenti. Per la valutazione della bontà del modello ci si basa su F1-Score a causa dello sbilanciamento esistente tra le due classi (si prevede ci siano molti più bambini non a rischio che a rischio, come conferma la percentuale stimata di incidenza, 5-33% [Overvelde, 2011, Reynolds, 2007]).

$$F1 - score = 2 \cdot \frac{Precisione \cdot Sensitività}{Precisione + Sensitività} \quad (6)$$

$$Precisione = \frac{TP}{TP + FP} \quad Sensitività = \frac{TP}{TP + FN} \quad (7 - 8)$$

	Predizione positiva	Predizione negativa
Campione positivo	TP True Positive	FN False Negative
Campione negativo	FP False Negative	TN True Negative

Tabella 3: matrice di confusione con classe predetta e classe reale.

Questo parametro di valutazione consiste nella media armonica tra *precisione* e *sensitività*, che rispettivamente rappresentano la porzione dei soggetti veri positivi (ossia correttamente classificati a rischio) tra tutti i soggetti predetti positivi e la porzione di soggetti positivi del test set che sono stati classificati come positivi. È stata posta particolare attenzione sulla sensitività coerentemente con la volontà di ottenere un modello capace di discriminare il maggior numero possibile di veri positivi, accettando l'errore di predire alcuni soggetti negativi come positivi. Il modello ideale è caratterizzato da F1-Score pari a uno e presenta quindi il massimo sia di sensitività che di precisione. Inoltre, si tiene in considerazione anche la media ponderata dell'F1-Score, ottenuta calcolando la media pesata (secondo la proporzione di ciascuna classe all'interno del train/test set) degli F1-Score relativi alla classe di interesse a rischio e di quella non a rischio.

3. RISULTATI

3.1 DESCRIZIONE DELLA POPOLAZIONE

Come sottolineato in precedenza, lo studio si è basato sui dati raccolti provenienti da 247 bambini, ma a causa di dropout e dati mancanti (missing), la somma delle informazioni non combacia sempre col numero totale dei soggetti.

Per quanto riguarda il genere, la popolazione era suddivisa in 128 maschi e 117 femmine (2 dati mancanti), mentre in riferimento alla mano dominante vi erano 211 destrimani, 25 mancini e 9 ambidestri (2 dati mancanti). Un'ulteriore informazione interessante è la lingua parlata dai partecipanti: 167 bambini sono di madrelingua italiana, mentre 72 sono di madrelingua straniera (8 dati mancanti). Tali dati sono rappresentati graficamente in Figura 7. Questo fattore in alcuni casi

potrebbe aver causato difficoltà nella comprensione delle indicazioni per lo svolgimento degli esercizi e di conseguenza cagionare incongruenza dei dati raccolti.

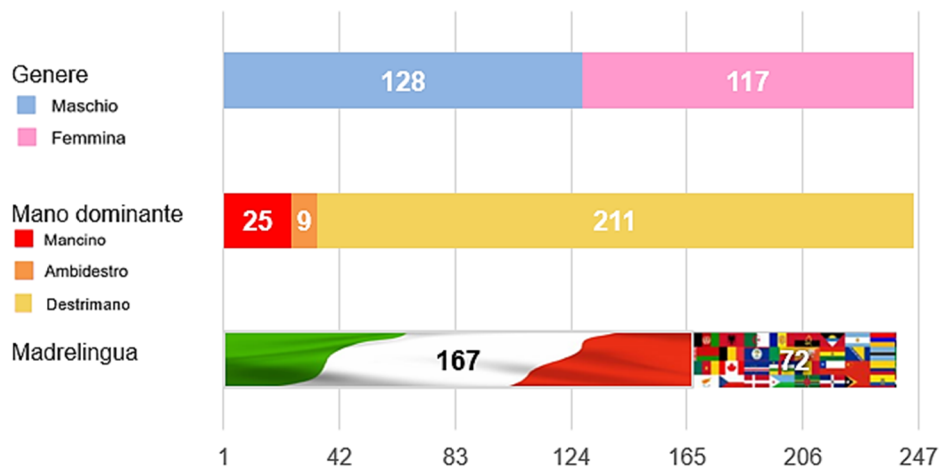
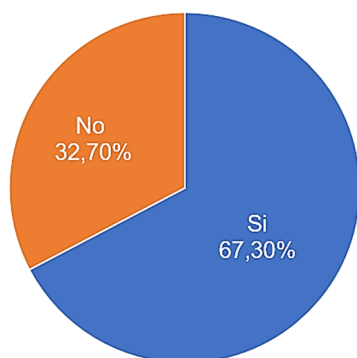
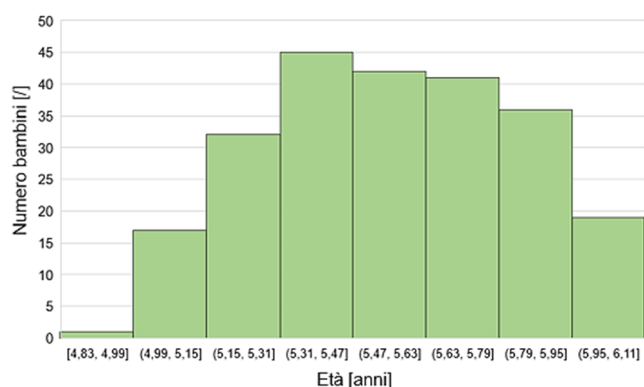


Figura 7: distinzione sui dati disponibili tra maschi e femmine in alto, mano dominante al centro e madrelingua dei soggetti in basso.

Altro dato interessante è la dimestichezza nell'utilizzo del tablet: 163 partecipanti lo avevano già utilizzato in precedenza, a differenza di 81 che non si erano mai interfacciati con il dispositivo (3 dati mancanti). Questo fattore potrebbe aver influito sulle prestazioni dei soggetti, in quanto un bambino abituato all'uso del tablet potrebbe dimostrare maggiore dimestichezza nella gestione e utilizzo della schermata interattiva. Infine, l'età dei soggetti, come già accennato in precedenza, è compresa tra i 4 e i 6 anni, con mediana di 5.58 anni e range interquartile di 0.42. I dati appena citati sono riportati rispettivamente in forma grafica in Figura 8a-8b.



(a)



(b)

Figura 8a-8b: percentuale sui dati disponibili di utilizzo del tablet a sinistra, distribuzione età sui dati disponibili dei soggetti a destra.

3.2 ANALISI DEI FATTORI CONFONDENTI

Dai test d'ipotesi utilizzati per l'analisi dei fattori confondenti (vedasi paragrafo 2.5) è emerso che i fattori confondenti che più influenzano le feature sono il genere, la mano dominante e la madrelingua.

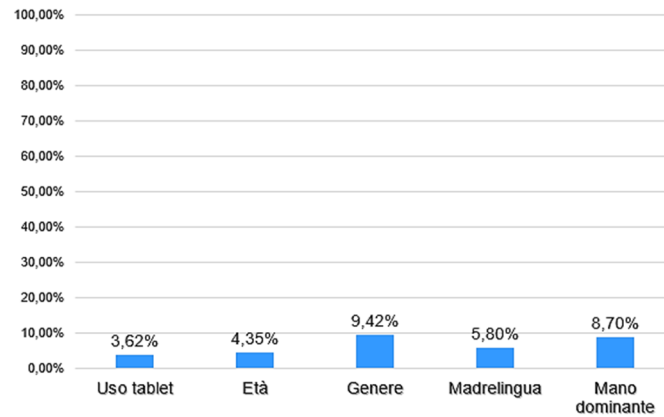


Figura 9: grafico a barre che indica la percentuale di feature influenzate da ciascun fattore confondente.

In particolare, il genere influenza il 9.42% delle feature, pari a 13 su 138, sotto riportate:

- *Centroide Y* nell'esercizio copia quadrato in condizione piccola;
- *Ampiezza della banda per la velocit * nell'esercizio copia quadrato in tutte le condizioni (spontanea, grande e piccola);
- *Energia del segnale per l'altitude* nell'esercizio copia quadrato in condizione piccola;
- *Energia del segnale per la velocit * nell'esercizio copia quadrato in condizione piccola;
- *Posizione primo picco per la pressione* nell'esercizio copia quadrato in condizione grande;
- *Velocit  massima* nell'esercizio tunnel quadrato;
- *Deviazione standard della velocit * nell'esercizio tunnel quadrato;
- *Ampiezza della banda per la velocit * nell'esercizio tunnel quadrato;
- *Ampiezza della banda per la velocit * nell'esercizio tunnel "ele";
- *Errore quadratico medio* nell'esercizio tunnel "ele".

La mano dominante influenza l'8.70% delle feature estratte, ovvero 12 su 138, di seguito riportate:

- *Varianza altitude* nell'esercizio copia quadrato in condizione grande;
- *Varianza azimuth* nell'esercizio copia quadrato in condizione grande;
- *Ampiezza della banda per l'altitude* nell'esercizio copia quadrato in condizione spontanea;
- *Ampiezza primo picco per l'azimuth* nell'esercizio copia quadrato in condizione grande;
- *Varianza azimuth* nell'esercizio tunnel quadrato;

- *Ampiezza della banda per l'azimuth* nell'esercizio tunnel quadrato;
- *Ampiezza e posizione primo picco per l'azimuth* nell'esercizio tunnel quadrato;
- *Energia per l'azimuth* nell'esercizio tunnel quadrato;
- *Varianza azimuth* nell'esercizio tunnel "ele";
- *Ampiezza primo picco per l'altitude* nell'esercizio tunnel "ele";
- *Coefficiente di correlazione R^2* nell'esercizio tunnel "ele".

Passando invece alla madrelingua, vengono influenzate 8 su 138 feature pari al 5.80%:

- *Velocità media* nell'esercizio copia quadrato per la condizione *grande*;
- *Energia per l'altitude* nell'esercizio copia quadrato per la condizione *grande*;
- *Ampiezza primo picco per l'azimuth* nell'esercizio tunnel quadrato;
- *Deviazione standard della velocità* nell'esercizio tunnel "ele";
- *Numero di picchi* nell'esercizio tunnel "ele";
- *Coefficiente di correlazione R^2* nell'esercizio tunnel "ele";
- *Numero errori* nell'esercizio tunnel "ele";
- *Numero esecuzioni* nell'esercizio tunnel "ele".

Col 4.35% l'età influenza 6 feature sulle 138 estratte, riportate di seguito:

- *In Air-time* nell'esercizio copia quadrato in condizione *grande*;
- *Posizione primo picco per la pressione* nell'esercizio copia quadrato in condizione *grande*;
- *Posizione primo picco per la velocità* nell'esercizio tunnel quadrato;
- *Posizione primo picco per la pressione* nell'esercizio tunnel "ele";
- *Coefficiente di correlazione R^2* nell'esercizio tunnel "ele";
- *Numero esecuzioni* nell'esercizio tunnel "ele".

Infine, in riferimento all'utilizzo del tablet sono influenzate 5 feature, pari al 3.62%:

- *Varianza azimuth* nell'esercizio copia quadrato in condizione *grande*;
- *Ampiezza della banda per l'altitude* nell'esercizio copia quadrato in condizione spontanea;
- *Ampiezza primo picco per l'azimuth* per l'esercizio copia quadrato in condizione *grande*;
- *Energia per l'altitude* per l'esercizio copia quadrato in condizione spontanea;
- *Numero di picchi* per l'esercizio copia quadrato.

Considerando solamente i fattori che influenzano più del 5% di feature – genere, mano dominante e madrelingua – risulta che i maschi classificati a rischio sono il 30%, mentre le femmine a rischio sono il 12.8% (Figura 10a). Osservando invece i risultati inerenti alla mano dominante, è stato stimato che i mancini a rischio sono il 32%, i destrimani a rischio il 19.5% e gli ambidestri a rischio il 55% (Figura 10c). Infine, i non madrelingua a rischio sono il 26.4%, mentre i madrelingua 20.4% (Figura 10b).

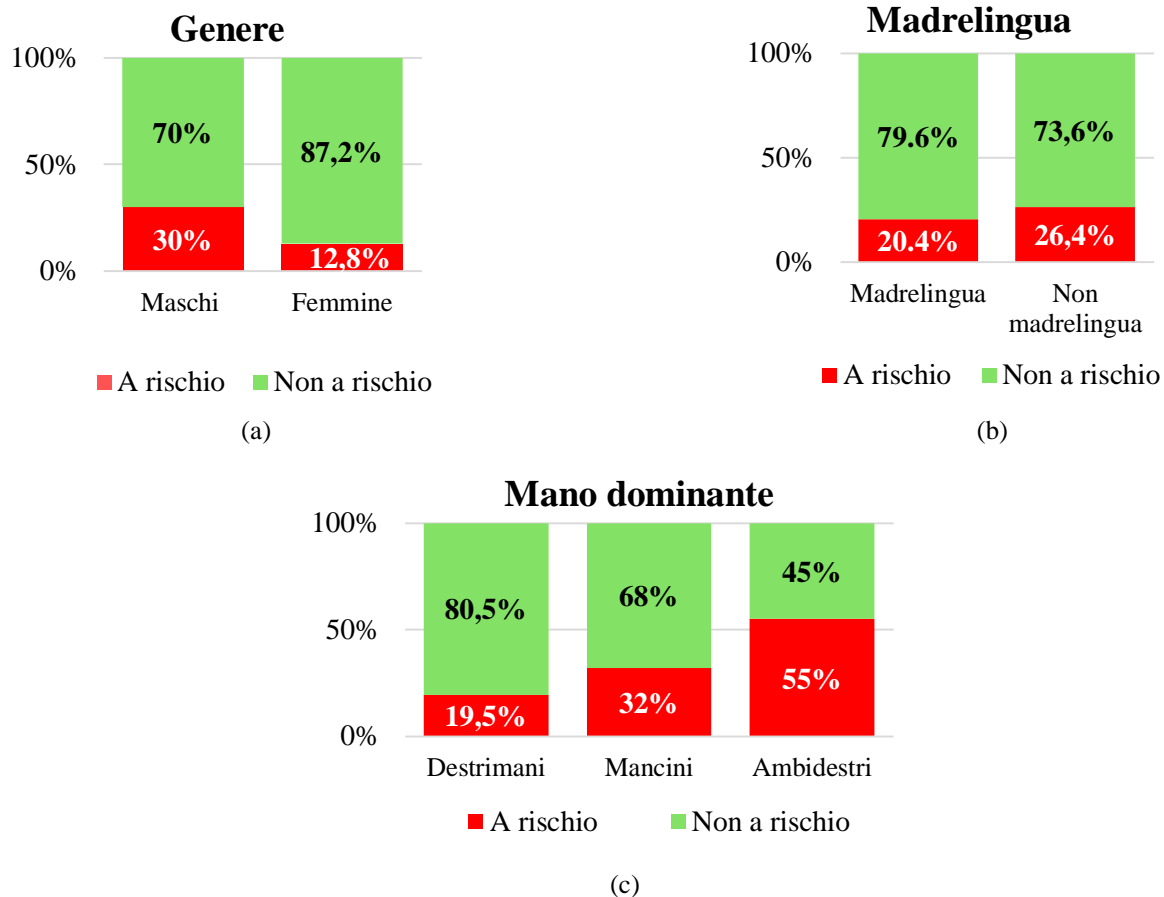


Figura 10: i grafici riportati mostrano i risultati relativi ai fattori confondenti che hanno avuto maggiore effetto sulle feature estratte: (a) genere, (b) madrelingua, (c) mano dominante.

3.3 IMPUTAZIONE E FEATURE ENGINEERING

L'imputazione è stata applicata a 17 delle 138 feature estratte, in particolare:

- 16 feature con percentuale di missing minore del 5%:
 - *Ampiezza banda per l'altitude* nell'esercizio copia quadrato in condizione spontanea;
 - *Ampiezza banda per l'altitude* nell'esercizio copia quadrato in condizione grande;

- *Ampiezza banda per l'azimuth* nell'esercizio copia quadrato in condizione grande;
 - *Ampiezza primo picco per l'altitude* nell'esercizio copia quadrato in condizione spontanea;
 - *Ampiezza primo picco per l'altitude* nell'esercizio copia quadrato in condizione grande;
 - *Ampiezza primo picco per l'azimuth* nell'esercizio copia quadrato in condizione grande;
 - *Posizione primo picco per l'altitude* nell'esercizio copia quadrato in condizione piccola;
 - *Posizione primo picco per l'altitude* nell'esercizio copia quadrato in condizione grande;
 - *Posizione primo picco per l'azimuth* nell'esercizio copia quadrato in condizione grande;
 - *Energia per l'altitude* nell'esercizio copia quadrato in condizione spontanea;
 - *Energia per l'altitude* nell'esercizio copia quadrato in condizione grande;
 - *Energia per l'azimuth* nell'esercizio copia quadrato in condizione grande;
 - *Ampiezza banda per l'atitude* nell'esercizio tunnel quadrato;
 - *Ampiezza primo picco per l'altitude* nell'esercizio tunnel quadrato;
 - *Posizione primo picco per l'altitude* nell'esercizio tunnel quadrato;
 - *Energia per l'altitude* nell'esercizio tunnel quadrato.
- 1 feature con percentuale di missing maggiore del 5%:
 - *Velocità nei tratti esterni al tunnel quadrato*, con l'utilizzo del modello di regressione SVM (Support Vector Machine), il quale presenta l'errore quadratico medio minore (RMSE: 1.9847) rispetto ai modelli di regressione addestrati.

Per quanto riguarda la feature engineering, sono state utilizzate per l'addestramento del modello due dataset, in uno le feature sono state trattate con normalizzazione e nell'altro con standardizzazione. Al momento dell'addestramento del modello predittivo e dell'estrazione delle metriche per la sua

valutazione si è constatato che la matrice di feature associata al modello migliore fosse quella trattata con normalizzazione.

3.4 CORRELAZIONE

La correlazione, eseguita nell'ambito della feature selection, è stata svolta al fine di stabilire eventuali relazioni lineari tra feature e la variabile rischio.

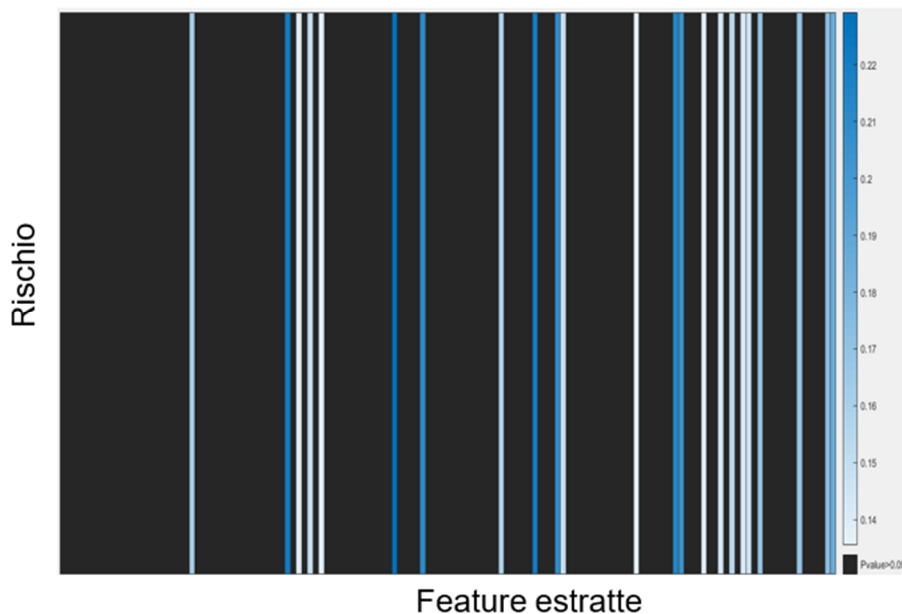


Figura 11: heatmap della correlazione tra feature e variabile rischio.

La Figura 11 si presenta colorata di nero laddove la correlazione tra feature e rischio risulti in un p-value > 0.05 , e dunque non significativa. Le feature che invece restituiscono una correlazione significativa presentano un p-value ≤ 0.05 .

In totale 23 feature su 138 mostrano una relazione lineare significativa con la variabile rischio:

- Per l'esercizio copia quadrato sono in air-time per la condizione *piccola* e nel campo delle frequenze: l'ampiezza di banda della velocità per la condizione *grande*, l'ampiezza del primo picco di Altitude per la condizione *spontanea* e *piccola*, l'ampiezza del primo picco di Azimuth per la condizione *grande*, l'energia di Azimuth per la condizione *piccola* e l'energia della velocità per la condizione *spontanea*;
- Per l'esercizio tunnel quadrato sono: la velocità media, il tempo sottosoglia, il Root-Mean-Square Error (RMSE) della regressione della Steering Law e nel campo delle frequenze l'energia della velocità, l'ampiezza di banda della pressione;

- Per l'esercizio tunnel "ele" sono: la velocità massima, la deviazione standard della velocità, il tempo sottosoglia, il coefficiente R2 della regressione della Steering Law, il numero di errori, il numero di esecuzioni con errori e nel campo delle frequenze l'ampiezza di banda di Azimuth, l'ampiezza del primo picco di Altitude, l'ampiezza del primo picco di pressione, l'ampiezza del primo picco di velocità, l'energia di Azimuth.

Successivamente è stata valutata la correlazione tra queste 23 feature, al fine di identificare eventuali collinearità tra di esse.

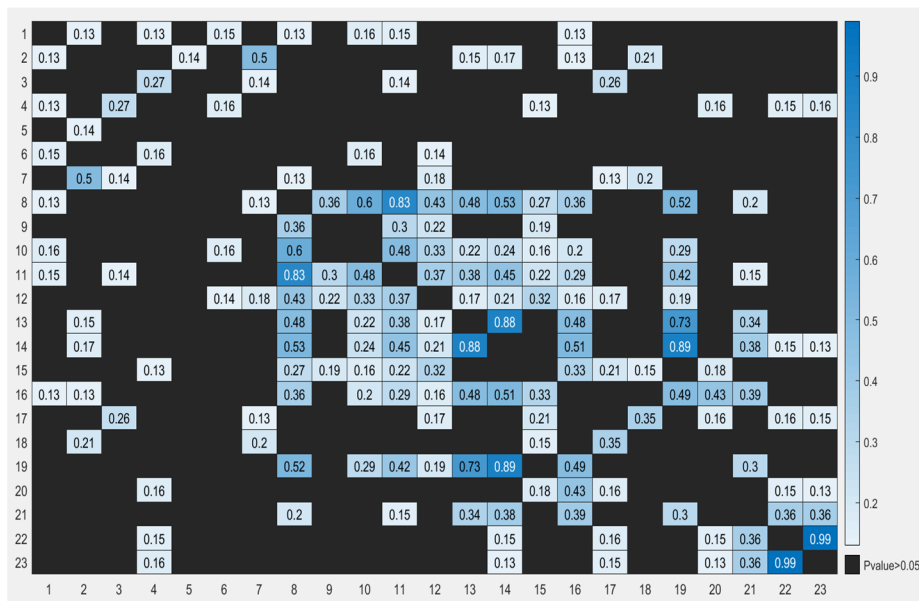


Figura 12: heatmap della correlazione tra feature.

La Figura 12 è colorata in nero laddove la correlazione tra le feature risulti in un p-value > 0.05 , mentre più la colorazione azzurra è intensa, maggiore è la correlazione tra le feature. Ciò significa che le due variabili con elevata correlazione non aggiungono maggiori informazioni se considerate entrambe e perciò è possibile predirne una dall'altra.

Nel nostro caso l'unica coppia di feature che ha correlazione maggiore di 0.9 è numero errori e numero esecuzioni (entrambe riferite all'esercizio tunnel "ele"). Al fine di eliminare la ridondanza di feature è stata scartata *numero esecuzioni*.

Dai grafici in Figura 13 è possibile vedere alcuni esempi di come le feature sopravvissute alla feature selection - utilizzate successivamente per addestrare il modello di classificazione - discriminino i due gruppi (bambini a rischio e bambini non a rischio di ritardi nell'apprendimento delle abilità grafiche) per quanto riguarda la media, oppure la forma della distribuzione o la presenza di outliers

(osservazioni con valore anomalo e chiaramente distante dall'insieme delle altre osservazioni disponibili).

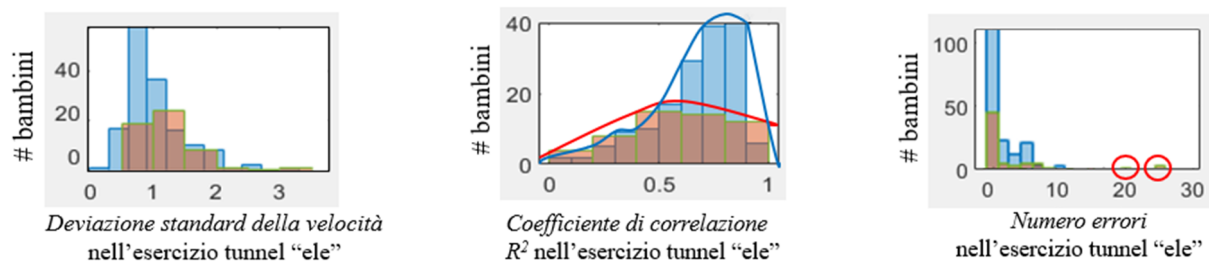


Figura 13: la Figura mostra alcuni esempi di discriminazione dei due gruppi di classificazione (in rosso i bambini a rischio e in blu quelli non a rischio) da parte di feature sopravvissute alla feature selection. Nel grafico a sinistra la media discrimina i due gruppi (infatti dal test di ipotesi risulta che la media della *Deviazione standard della velocità* nell'esercizio tunnel "ele" per il gruppo dei bambini a rischio - 1.22 - e del gruppo non a rischio - 1.02 - sono significativamente differenti); nel grafico al centro si possono notare le differenti forme di distribuzione del *Coefficiente di correlazione R^2* nell'esercizio tunnel "ele" per i due gruppi; infine, nel grafico a destra si evidenzia attraverso i cerchietti rossi la presenza di outliers nel gruppo a rischio per quanto riguarda *Numero errori* nell'esercizio tunnel "ele".

3.5 ADDESTRAMENTO E VALUTAZIONE DEL MODELLO

Si valuta ora nello specifico la fase di addestramento del modello. Il nostro dataset è composto da 247 bambini, ma ne sono stati presi in analisi solo 212 poiché dei rimanenti non si era in possesso della classificazione delle insegnanti e alcuni non sono stati presi in considerazione a causa di problemi di acquisizione. Dalle informazioni a disposizione risultava che il 25% dei bambini era classificato a rischio di ritardi nell'apprendimento delle abilità grafiche, come risulta dalla letteratura. Sono stati assegnati 2/3 dei bambini a rischio al train set e 1/3 dei bambini a rischio al test set in modo da avere i due set stratificati; lo stesso è stato fatto per i bambini classificati come non a rischio. Si sono ottenuti così i seguenti set: nel train set sono stati collocati 36 bambini a rischio e 106 non a rischio, nel test set 17 a rischio mentre 53 non a rischio.

Fissata la PCA con il 95% di varianza spiegata (15 risultano essere le componenti principali sufficienti, Figura 14) e abilitata la tecnica *10-fold-crossvalidation* sono stati addestrati tutti i modelli singolarmente sui dati forniti al software.

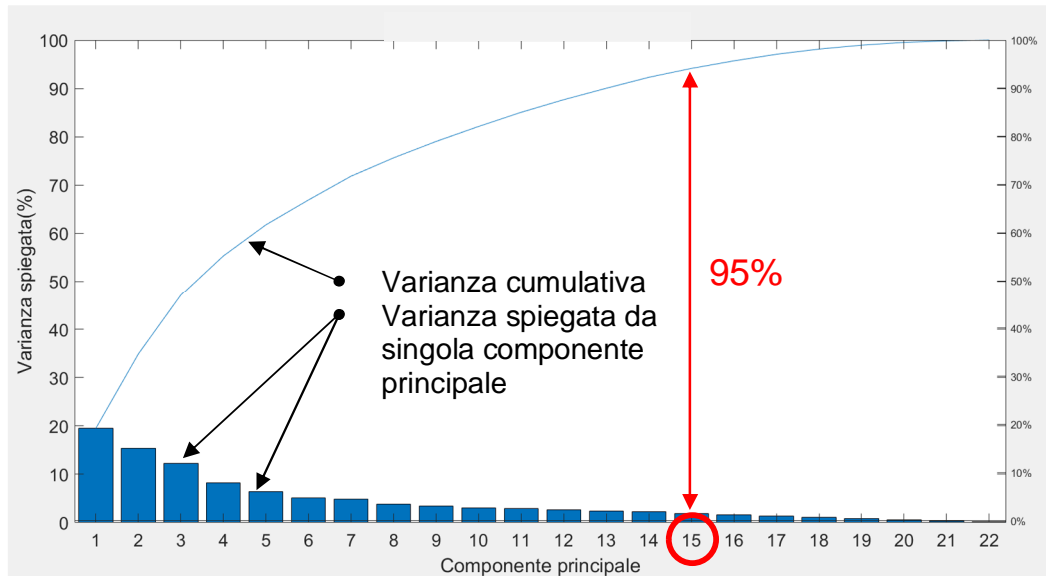


Figura 14: screeplot della varianza spiegata dalle singole componenti principali e quella cumulativa.

Per la scelta del modello si è fatto riferimento, come detto in precedenza, al valore assunto dall' F1-Score ottenuto considerando come classe di interesse i bambini classificati come a rischio; quindi sono stati eliminati i modelli aventi tale valore minore di 0.4. Sulla base di ciò è stato selezionato come modello migliore il *RUSBoosted trees* poiché mostrava F1-Score maggiore pari a 0.553 (Figura 15).

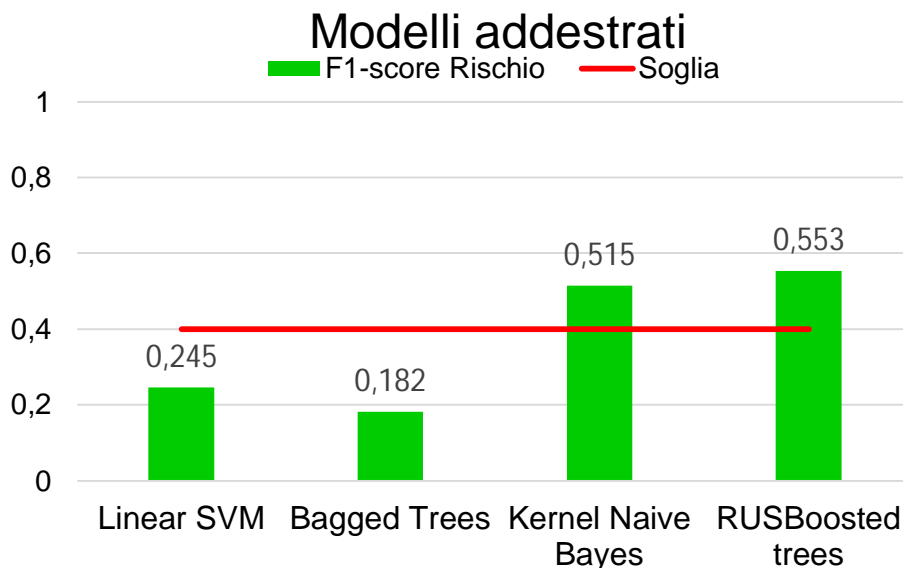


Figura 15: in Figura sono stati riportati gli F1-Score dei 4 modelli che hanno prestazioni migliori tra quelli addestrati.

Inoltre, tale modello mostra una sensibilità per il rischio pari a 0.61.

La percentuale di miglioramento nell'individuazione della classe di interesse (a rischio) ottenuta utilizzando il modello scelto anziché uno no skill (ovvero un modello casuale senza vincoli sulle metriche) è risultata essere pari a circa 25%. Per il calcolo di tale metrica è stata utilizzata la formula 12 data dal rapporto che ha come numeratore la differenza tra l'F1-Score del modello scelto e dell'F1-Score del modello pessimo e come denominatore l'F1-Score del modello pessimo, dove per modello pessimo si intende quello la cui precisione risulta minima, ovvero non in grado di identificare alcun soggetto classificato come a rischio.

$$\% \text{ miglioramento} = \frac{F1Score \text{ modello scelto} - F1Score \text{ modello pessimo}}{F1Score \text{ modello pessimo}} \cdot 100 \quad (9)$$

Infine, si è calcolato la media ponderata dell'F1-Score che risulta pari a 0.71 al fine di tenere conto, nella scelta del modello, oltre che alla capacità dei modelli di individuare la classe a rischio, anche a quella di distinguere coloro che non sono a rischio. Successivamente si è proseguito con la valutazione del modello: in questo caso è stato applicato ai dati di test.

È emerso un valore di sensitività pari a 0.59 e un F1- score di 0.49 nel caso in cui la classe di interesse sia il rischio, mentre in caso di media ponderata tra F1-Score considerando come classe di interesse rischio e F1-Score considerando come classe di interesse i bambini classificati come non a rischio è pari a 0.71. Il miglioramento rispetto all'utilizzo di un modello no skill è pari a 24.8%.

Si riporta qui di seguito – Figura 16 – la matrice di confusione in cui si mostra la classe predetta e la classe reale, da cui emerge che il modello ha predetto 10 soggetti a rischio (veri positivi) su 17 e 39 soggetti non a rischio (veri negativi) su 53. Il numero dei falsi positivi è pari a 14, mentre il numero dei falsi negativi è pari a 7.

Classe Reale	Non rischio	39	14
	Rischio	7	10
		Non Rischio	Rischio
		Classe Predetta	

Figura 16: matrice di confusione test set.

4. DISCUSSIONE

Riprendendo quanto citato in precedenza, la disgrafia fa parte dei Disturbi Specifici dell'Apprendimento (DSA) e comporta difficoltà nell'espressione scritta, cioè irregolarità per quanto riguarda la forma e la grandezza delle lettere, la spaziatura tra le lettere e tra le parole, lo spelling, la grammatica, la punteggiatura, l'impugnatura della penna, la velocità di scrittura e l'organizzazione spaziale [American Psychiatric Association, 2013]. Ad oggi si ritiene che una diagnosi corretta della disgrafia non possa essere effettuata prima della fine del secondo anno di scuola primaria. Il nostro studio ha come obiettivo principale la possibilità di poter valutare il rischio di ritardi nell'apprendimento delle abilità grafiche fin dal pregrafismo, creando un modello predittivo che possa individuare tale rischio in bambini che frequentano l'ultimo anno di Scuola dell'Infanzia. In aggiunta seguono poi anche i due obiettivi secondari:

1. Analisi dei fattori confondenti;
2. Identificazione delle feature più predittive.

Tutti gli obiettivi sopra elencati sono stati portati a termine e di seguito nei paragrafi successivi verranno trattati singolarmente.

4.1 ADDESTRAMENTO E VALUTAZIONE DEL MODELLO

Venendo ora alle caratteristiche del modello scelto, come learner type il RUSBoosted trees fa parte della categoria decision tree, che dimostra prestazioni maggiori rispetto agli altri type. Tale modello, infatti, consiste in un algoritmo che generalmente viene applicato nelle classificazioni binarie in cui i dati sono sbilanciati, esattamente come nel caso in esame; in aggiunta ha la peculiarità di aumentare i pesi per le osservazioni classificate in modo non corretto ad ogni step e, quindi, di riclassificare in modo corretto i dati che erano identificati in modo sbagliato allo step precedente.

Nella fase di addestramento le prestazioni di tale modello risultano piuttosto simili a quelle del modello Kernel Naive Bayes, come si può vedere in Figura 14, ma è stato scelto il RUSBoosted trees coerentemente alla volontà di privilegiare, in primis, il modello che avesse F1-Score maggiore; in secondo luogo, si è ricercata la media ponderata degli F1-Score piuttosto elevata al fine di ottenere un modello che mostrasse una buona capacità sia nell'individuare la classe a rischio, ma anche, non di minor importanza, nel distinguere coloro che non sono a rischio. Questa seconda condizione si verifica per entrambi i modelli e si presenta pari a 0.722 per il modello RUSBoosted trees e pari a

0.768 per il modello Kernel Naive Bayes. Infine, la sensitività per la classe a rischio è maggiore per il primo modello (pari a 0.722) rispetto a quella del secondo modello (pari a 0.472).

Anche durante la fase di valutazione il modello RUSBoosted trees ha dimostrato prestazioni maggiori rispetto alla soglia prefissata (pari a 0.4), con un F1-Score pari a 0.49. Inoltre, il miglioramento rispetto all'utilizzo di un modello no skill, pari a 24.8%, risulta molto simile a quello calcolato durante la fase di train, pari a 25%, indice del fatto che il modello generalizza bene su dati nuovi.

Infine, osservando la matrice di confusione (Figura 16) si profila un'altra caratteristica del modello: il numero di falsi positivi (14) è maggiore rispetto a quello dei falsi negativi (7), coerentemente alla volontà di ottenere un modello che nei casi incerti individuasse dei falsi positivi a cui un percorso di potenziamento potrebbe migliorare le eventuali carenze nell'abilità di scrittura, seppur in assenza di un effettivo rischio di ritardo nell'apprendimento delle abilità grafiche.

4.2 ANALISI FATTORI CONFONDENTI

In riferimento al grafico in Figura 9, relativo ai risultati ottenuti dall'analisi dei fattori confondenti, si osserva che, in generale, i potenziali fattori confondenti non hanno avuto grande effetto sulla maggior parte della feature estratte.

Considerando significativi, come detto in precedenza, solamente quelli che hanno raggiunto il 5% di influenza si trovano risultati coerenti a quanto emerge dalla letteratura circa il genere e la mano dominante: maschi e mancini hanno maggiore incidenza di disgrafia [Rubin et al, 1982; Rosenblum et al, 2004]. In particolare, i mancini a rischio sono in numero maggiore rispetto ai destrimani, a conferma di ciò che si trova in letteratura, ma inferiore agli ambidestri; ciò potrebbe essere dovuto al fatto che la maggior parte dei bambini identificati come ambidestri abbiano dimostrato incertezza nella scelta della mano dominante e, di conseguenza, non abbiano ancora padronanza nella prensione e nell'utilizzo della penna (infatti la mano dominante influenza feature inerenti ad altitude e ad azimuth sia nel dominio del tempo che delle frequenze).

In aggiunta, dallo studio condotto, si può fare un'ulteriore considerazione su madrelingua: questo fattore ha avuto effetto sul 6% delle feature, alcune delle quali mostrano una relazione con esso, come ad esempio numero errori (feature che conta il numero degli errori commessi durante l'esecuzione dell'esercizio), che presenta un valore elevato nel caso dei non madrelingua, probabilmente a causa del fatto che questi non capivano esattamente il compito a loro assegnato; mentre per altre feature influenzate da madrelingua, come ad esempio quelle inerenti all'altitude, non sembra esserci una spiegazione valida.

Per quanto riguarda l'età non ci sono differenze significative tra i vari soggetti: in letteratura [Overvelde, 2011] vengono confrontati bambini di classi diverse con differente numero di anni di istruzione durante i quali migliorano le loro capacità di scrittura, mentre nel caso in esame i dati riguardano bambini la cui età si differenzia solo in termini di mesi, che quindi non determina significative discrepanze nel grado di sviluppo delle abilità grafiche.

Infine, la dimestichezza con il tablet risulta avere influenza sulle feature estratte da altitude e azimuth. Si ipotizza che un bambino con esperienza nell'utilizzo del tablet o della penna per scrivere su tablet possa essere in grado di impugnare meglio la penna. Questa affermazione però va consolidata da ulteriori esperimenti: infatti i bambini iniziano a decidere come impugnare la penna proprio in questa fase di età, quindi sarebbe stato un lavoro più completo se si fosse potuto raccogliere i dati su due archi temporali differenti, per confrontare eventuali cambiamenti.

Si sottolinea il fatto che le percentuali d'influenza trovate sono molto basse e, inoltre, dopo le tecniche di feature selection le feature influenzate da più fattori confondenti sono state eliminate, quindi non è stata presa alcuna contromisura nel modello finale.

4.3 ANALISI FEATURE PIÙ PREDITTIVE

Attraverso la feature selection è stato possibile stabilire quali feature fossero maggiormente correlate alla variabile di classificazione ed eliminare la ridondanza di informazioni apportate. In particolare, la scelta di mantenere *Numero errori* nell'esercizio tunnel "ele" - invece che *Numero esecuzioni* nell'esercizio tunnel "ele" - dalla coppia di feature con alta correlazione riscontrata è stata presa perché si ritiene che questa feature quantifichi in modo migliore il numero complessivo di errori commessi nelle prove dell'esercizio tunnel "ele".

In aggiunta, dal nostro studio emerge che le feature più predittive sono quelle legate ad altitude, azimuth e velocità. Dalla letteratura emergono risultati discordanti riguardo alle variabili maggiormente capaci di identificare il soggetto disgrafico: in alcuni studi risultano essere più predittive le variabili legate alla cinematica – ad esempio in Asselborn et al, 2018 –, in altri le caratteristiche riferite all'inclinazione della penna – in Mekyska et al, 2017 – e in altri ancora quelle inerenti alla pressione – come in Asselborn et al, 2020 -. Possiamo concludere di essere in accordo con i risultati degli studi di Asselborn et al, 2018 e Mekyska et al, 2017, pur tenendo conto che il nostro studio si concentra sul pregrafismo mentre quelli sopra citati sulla scrittura.

4.4 CONCLUSIONI

4.4.1 LIMITI DELL'ANALISI

All that glisters is not gold, diceva William Shakespeare ne “Il mercante di Venezia”, allo stesso modo potremmo dire della nostra analisi, difatti di seguito sono riportati i limiti attuali del nostro lavoro che, tramite analisi future, potranno essere minimizzati.

Innanzitutto, il numero di campioni a disposizione – 247 soggetti – potrebbe non essere sufficiente allo scopo di previsioni basate su dati raccolti da soggetti in età precoce. Inoltre, l'impossibilità di ricevere dati mancanti di alcuni soggetti ha diminuito ulteriormente la quantità effettiva di informazioni a nostra disposizione. Un altro aspetto limitante è dovuto al fatto che nella fase di addestramento del modello – vedasi paragrafo 2.4 – è stato preso come riferimento il giudizio delle maestre, che non può garantire certezza oggettiva in quanto non proveniente da uno specialista. In aggiunta, l'evoluzione rapida delle abilità grafiche nella fascia d'età considerata può compromettere la classificazione finale, in quanto non è stato possibile valutarne l'andamento in tempi brevi e quindi il miglioramento nel tempo. Infine, facendo riferimento a quanto detto nel paragrafo 2.2, la nostra analisi si è limitata allo studio di solo tre dei molteplici esercizi compiuti dai soggetti rispetto a tutti quelli fruibili tramite l'applicazione *Play Draw Write*.

In definitiva, per ottenere studi maggiormente approfonditi un possibile sviluppo futuro potrebbe essere quello di incrementare il numero di soggetti presi in esame e ampliare lo spettro dell'analisi anche agli altri esercizi non valutati.

4.4.2 SVILUPPI FUTURI

Dal modello realizzato si ottiene una classificazione di quali bambini sono a rischio di ritardi nell'apprendimento delle abilità grafiche o meno, basandosi solo su simboli pregrafici e senza ricorrere all'uso della scrittura. I bambini classificati come a rischio verranno sottoposti ad un percorso di potenziamento che consisterà in attività mirate a sviluppare i prerequisiti della scrittura, e in seguito tali soggetti saranno poi rimonitorati a distanza di anni per verificare che le attività da loro svolte abbiano portato ad un miglioramento consistente.

4.4.3 CONCLUSIONE FINALE

Tra le novità apportate rispetto alla letteratura è importante sottolineare che il modello predittivo creato è in grado di individuare le problematiche del gesto grafico e di predire in modo oggettivo il rischio di ritardi nello sviluppo di abilità grafiche fin dal pregrafismo. Collegandoci a quest'ultimo aspetto è interessante il fatto che il modello si basa sull'analisi di esercizi pregrafici, togliendo dunque il vincolo della capacità di scrittura per i soggetti presi in considerazione, e potendo così valutare un eventuale rischio di ritardi nell'apprendimento delle abilità grafiche fin dalla scuola dell'Infanzia.

In conclusione, riteniamo che la nostra analisi possa essere un ottimo punto di partenza per studi futuri.

BIBLIOGRAFIA

1. J. Accot, S. Zhai, 1997, “Beyond Fitt’s Law: Models for trajectory-based HCI tasks”, *Proceedings of ACM CHI’97 Conference on Human Factors in Computing Systems*, 295-302;
2. American Psychiatric Association, 2013, “Diagnostic and statistical manual of mental disorders”, 5th ed., *Washington, DC: American Psychiatric Association*;
3. T. Asselborn, T. Gargot, Ł. Kidziński, W. Johal, D. Cohen, C. Jolly, P. Dillenbourget, 2018, “Automated human-level diagnosis of dysgraphia using a consumer tablet”, *npj Digit Medicine*, 1, 42;
4. T. Asselborn, M. Chapatte, P. Dillenbourg, 2020, “Extending the Spectrum of Dysgraphia: A Data Driven Strategy to estimate Handwriting Quality”, *Nature Medicine*;
5. A. Binet, J. Courtier, 1893, “Sur la vitesse des mouvements graphiques”, *Revue Philosophique de la France et de l’Étranger*, 664–671;
6. P. J. Chung, D. R. Patel, I. Nizami, 2020, “Disorder of written expression and dysgraphia: definition, diagnosis, and management”, *Transl Pediatr*, 9(Suppl 1):S46-S54;
7. J. Danna, J.-L. Velay, V. Paz-Villagràn, A. Capel, C. Petroz, C. Gondre, et al, 2013, “Handwriting movement sonification for the rehabilitation of dysgraphia”, *10th International Symposium on Computer Music Multidisciplinary Research, Marseille, October 15–18* (accepted);
8. R. K. Deuel, 1995, “Developmental dysgraphia and motor skills disorders”, *J Child Neurol*, 10 Suppl 1:S6-8;
9. T. H. Falk, C. Tam, H. Schellnus, T. Chau, 2011, “On the development of a computer-based handwriting assessment tool to objectively quantify handwriting proficiency in children”, *Computer Methods and Programs in Biomedicine*, 104(3), 102–111.
10. W. T. Fitch, M. D. Martins, 2014, “Hierarchical processing in music, language, and action: Lashley revisited”, *Ann. N. Y. Acad. Sci.* 1316, 87–104;
11. K. P. Feder, A. Majnemer, 2007, “Handwriting development, competency, and intervention”, *Dev. Med. Child Neurol* 49, 312–317;
12. M. A. Hall, L. A. Smith, 1998, “Practical feature subset selection for machine learning. In C.” *McDonald(Ed.), Computer Science ’98 Proceedings of the 21st Australasian Computer Science Conference ACSC’98, Perth, 4-6 February*, pp. 181-191;
13. L. Hamstra-Bletz, A. W. Blöte, 1993, “A longitudinal study on dysgraphic handwriting in primary school”, *Journal of Learning Disabilities*, 26, 689-699;

14. L. Hamstra-Bletz, J. de Bie, B. den Brinker, 1987, "Beknopte beoordelingsmethode voor kinderhandschriften: BHK – Concise Evaluation Scale for Children's Handwriting: BHK. Lisse", *Swets and Zeitlinger*;
15. R. Karlsdottir, T. Stefansson, 2002, "Problems in developing functional handwriting", *Percept Mot Skills*, 94: 623–662;
16. K. S. Lashley, 1951, "The problem of serial order in behavior", *Bobbs-Merrill*, 112–136;
17. F. Lunardini, S. Maggioni, C. Casellato, M. Bertucco, A. L. G. Pedrocchi, T. D. Sanger, 2015b, "Increased task-uncorrelated muscle activity in childhood dystonia", *Journal of NeuroEngineering and Rehabilitation*, 12, 52;
18. G. R. Lyon, N. A. Krasnegor, 1996, "Attention, memory, and executive function", *Paul H Brookes Publishing Co.*;
19. J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, S. Rosenblum, 2017, "Identification and rating of developmental dysgraphia by handwriting analysis", *IEEE Trans. Hum.–Mach. Syst.* 47, 235–248;
20. A. Overvelde, W. Hulstijn, 2011, "Handwriting development in grade 2 and grade 3 primary school children with normal, at risk, or dysgraphic characteristics", *Res Dev Disabil*, 32:540-8;
21. E. Pagliarini, M. T. Guasti, C. Toneatto, E. Granocchio, F. Riva, D. Sarti, B. Molteni, N. Stucchi, 2015, "Dyslexic children fail to comply with the rhythmic constraints of handwriting", *Human Movement Science*, 42, 161-182;
22. C. Reynolds, 2007, "Encyclopedia of special education: A reference for the education of children, adolescents, and adults with disabilities and other exceptional individuals", *New York, NY: John Wiley & Sons*, 3rd ed;
23. S. Rosenblum, S. Parush, P. L. Weiss, 2003, "Computerized temporal handwriting characteristics of proficient and non-proficient handwriters", *The American Journal of Occupational Therapy*, 57, 129-138;
24. S. Rosenblum, S. Parush, P. L. Weiss, 2004, "Handwriting evaluation for developmental dysgraphia: Processes versus product.", *Reading and writing*, 17 (5), pp. 433-458;
25. N. Rubin, S.E. Henderson, 1982, "Two sides of the same coin: Variation in teaching methods and failure to learn to write Special Education", *Forward Trends*, 9 (4) (1982), pp. 17-24;
26. B. Smits-Engelsman, A. Niemeijer, G. van Galen, 2001, "Fine motor deficiencies in children diagnosed as DCD based on poor graphomotor ability", *Hum. Mov. Sci.* 20, 161–182.
27. R. H. Stetson, J. A. McDill, 1923, "Mechanism of the different types of movement", *Psychol. Monogr*, 32, 18–40;

28. H. Teulings, F. J. Maarse, 1984, "Digital recording and processing of handwriting movements", *Human Movement Science*, Volume 3, Issues 1–2, 1984;
29. G. P. Van Galen, W. P. De Jong, 1995, "Fitts' law as the outcome of a dynamic noise filtering model of motor control", *Nijmegen Institute for Cognition and Information (NICI)*, 14, 4-5, 539-571;
30. P. Van Galen, H. L. Teulings, 1983, "The independent monitoring of form and scale factors in handwriting", *Acta Psychol*, (Amst) 54, 9–22;
31. P. Viviani, C. Terzuolo, 1982, "Trajectory determines movement dynamics", *Neuroscience*, 7(2), 431–437;