

Classificazione del discorso immaginato usando il siamese Rete neurale

Dong-Yeon Lee

Dept. Brain and Cognitive Engineering Korea
University Seoul, Repubblica di
Corea dongyeon_lee@korea.ac.kr

Minji Lee

Dept. Brain and Cognitive Engineering Korea
University Seoul, Repubblica di
Corea minjilee@korea.ac.kr

Seong-Whan Lee

Dept. Artificial Intelligence
Dept. Brain and Cognitive Engineering Korea
University Seoul, Repubblica di
Corea sw.lee@korea.ac.kr

Riassunto : il discorso immaginato è messo in luce come una nuova tendenza nell'interfaccia cervello-macchina grazie alla sua applicazione come strumento di comunicazione intuitivo. Tuttavia, studi precedenti hanno mostrato prestazioni di classificazione basse, pertanto il suo utilizzo nella vita reale non è fattibile. Inoltre, non è stato trovato alcun metodo adatto per analizzarlo. Recentemente, gli algoritmi di deep learning sono stati applicati a questo paradigma. Tuttavia, a causa della scarsa quantità di dati, l'aumento delle prestazioni di classificazione è limitato. Per affrontare questi problemi, in questo studio, abbiamo proposto un framework end-to-end utilizzando il codificatore di rete neurale siamese, che apprende le caratteristiche discriminanti considerando la distanza tra le classi. Le parole immaginate (ad esempio, arriba (su), abajo (giù), derecha (destra), izquierda (sinistra), adelante (avanti) e atrás (indietro)) sono state classificate utilizzando i segnali grezzi dell'elettroencefalografia (EEG). Abbiamo ottenuto una precisione di classificazione di 6 classi del 31.40 ± 2.73% per il discorso immaginario, che è stato possibile perché è stata utilizzata la rete neurale siamese, che aumenta la distanza tra campioni dissimili diminuendo la distanza tra campioni simili. A questo proposito, il nostro metodo può apprendere caratteristiche discriminanti da un piccolo set di dati. Il quadro proposto aiuterebbe ad aumentare le prestazioni di classificazione del discorso immaginato per una piccola quantità di dati e ad implementare un sistema di comunicazione intuitivo.

Parole chiave: discorso immaginario, struttura end-to-end, siamese rete neurale, deep learning, interfaccia cervello-macchina

I. INTRODUZIONE

L'interfaccia cervello-macchina (BMI) ha molteplici applicazioni come il controllo di dispositivi esterni, sistemi di comunicazione e altro; per i quali utilizza segnali cerebrali. Tradizionalmente questa interfaccia è stata utilizzata da persone gravemente disabili, tuttavia, recentemente viene utilizzata da persone sane per la comunicazione o per i sistemi di assistenza alla vita quotidiana [1-3]. Tra questi, l'elettroencefalografia

(EEG) è comunemente usato per il suo basso costo e convenienza nel campo del BMI [4, 5]. Sotto il paradigma BMI, ci sono due modi per generare segnali cerebrali [6]. Uno è paradigmi esogeni e l'altro è paradigmi endogeni. I paradigmi esogeni richiedono la presentazione di uno stimolo esterno come la stimolazione visiva tremolante, la stimolazione uditiva e altri segnali cerebrali evocati [7-9]. D'altra parte, i paradigmi endogeni non richiedono uno stimolo esterno [10]; i soggetti eseguono un compito mentale (cioè, immagini motorie (MI), immagini vive e discorso immaginario) mentre vengono registrati i segnali cerebrali [11, 12]. Il paradigma endogeno può riflettere l'intenzione dell'utente ed essere applicato a sistemi più intuitivi, pertanto la sua applicazione nella tecnologia BMI è in aumento.

Il paradigma più utilizzato nel BMI endogeno convenzionale è MI. Nei paradigmi MI, i segnali cerebrali sono generati attraverso l'immaginazione dell'intenzione di movimento [10].

È comunemente usato per controllare dispositivi protesici come i bracci robotici. Tuttavia, il numero di classi IM immaginabili è limitato (cioè, mano sinistra, mano destra, lingua e piede). Quando si utilizza l'MI per sistemi di comunicazione come gli ortografi [13], non può essere un sistema intuitivo; perché l'utente immagina non la parola che vuole inserire ma un'azione specifica [14, 15]. Pertanto, un paradigma intuitivo per risolvere questi problemi ha iniziato a venire alla ribalta.

Il discorso immaginato è una nuova tendenza del paradigma BMI che conduce la pronuncia interna delle parole senza alcun movimento e output udibile [16]. Questo paradigma fino ad ora ha utilizzato il tradizionale metodo di estrazione e classificazione delle caratteristiche del BMI [14, 15, 17]. Il metodo di estrazione delle caratteristiche comunemente usato è il modello spaziale comune (CSP) [14, 17]. In Lee et al. [14], hanno classificato 13 classi di discorso immaginario utilizzando il CSP come metodo di estrazione delle caratteristiche. Per i metodi di classificazione sono state utilizzate l'analisi discriminante lineare regolarizzata del restringimento (RLDA) e la foresta casuale (RF). Di conseguenza, RF ha mostrato la massima precisione; 20,4% per la classificazione a 13 classi (12 parole immaginate con lo stato di riposo). In Dasalla et al. [17], CSP e support vector machine (SVM) sono stati applicati per la classificazione delle immagini vocaliche binarie. Le accuratèzze di classificazione complessive variavano dal 68% al 78%. CSP ha il vantaggio di ridurre il segnale ad alta dimensionalità nel segnale a bassa dimensionalità e può massimizzare la distanza tra classi diverse. Tuttavia, CSP è ottimizzato per problemi di classi binarie [18]. Pertanto, sono stati utilizzati anche altri metodi per l'estrazione

Questo lavoro è stato sostenuto in parte dalla sovvenzione dell'Institute for Information & Communications Technology Promotion (IITP), finanziata dal governo della Corea (MSIT) (n. 2015-0-00185, Development of Intelligent Pattern Recognition Softwares for Ambulator Brain Computer Interface; n. 2017-0-00451, Sviluppo di una tecnologia di elaborazione cerebrale e cognitiva basata su BCI per il riconoscimento delle intenzioni dell'utente mediante il deep learning, n.

© 20xx IEEE. È consentito l'uso personale di questo materiale. È necessario ottenere l'autorizzazione da IEEE per tutti gli altri usi, in qualsiasi media attuale o futuro, inclusa la ristampa/ripubblicazione di questo materiale per scopi pubblicitari o promozionali, la creazione di nuove opere collettive, per la rivendita o la ridistribuzione a server o elenchi o il riutilizzo di qualsiasi componente protetto da copyright di quest'opera in altre opere.

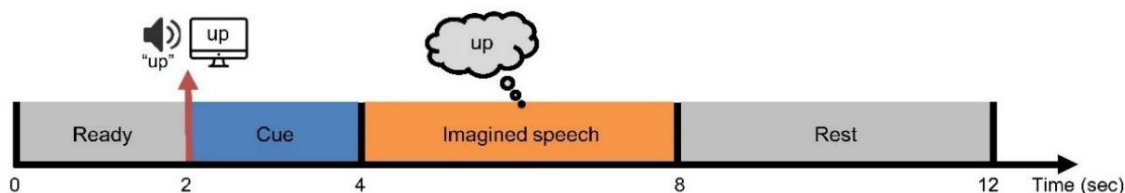


Fig. 1. Procedura sperimentale del discorso immaginato.

[19], la classificazione di 5 vocali e 6 parole era mirata, hanno usato la trasformata wavelet discreta (DWT) per l'estrazione delle caratteristiche e RF per la classificazione. Questi metodi hanno mostrato prestazioni vicine al livello casuale (vocale: 22,72%, parole: 19,60%). In Garcia-Salinas et al. [20], hanno utilizzato un sacchetto di caratteristiche (BoF) per l'estrazione delle caratteristiche e Naive Bayes per la classificazione. Hanno mostrato una precisione media del 65,65% nella classificazione delle parole immaginate in 5 classi. Tuttavia, hanno affermato che le prestazioni erano elevate perché la presentazione degli stimoli durante il processo di raccolta dei dati non era randomizzata. Fino ad ora, questi metodi di estrazione e classificazione delle caratteristiche non hanno ancora mostrato prestazioni ottimali nel paradigma del discorso immaginato [21]. Pertanto, le prestazioni di classificazione del parlato immaginato possono essere migliorate selezionando caratteristiche e classificatori appropriati [22]. Tuttavia, non è facile trovare un metodo di analisi adatto per il discorso immaginato.

Per affrontare questi problemi e migliorare le prestazioni, recentemente sono aumentati gli studi sull'analisi del linguaggio immaginato attraverso il deep learning. Un metodo di apprendimento profondo comunemente utilizzato è la rete neurale convoluzionale (CNN). In Saha et al. [23], per classificare la classificazione binaria delle categorie fonologiche, hanno utilizzato la covarianza incrociata del canale per la preelaborazione. Il framework del modello era composto da CNN, rete a lungo-breve termine (LSTM) e un codificatore automatico profondo per estrarre le informazioni spaziotemporali. Hanno raggiunto la migliore precisione media del 77,9% in cinque diverse attività di classificazione binaria. Nel loro recente lavoro [24], hanno utilizzato la covarianza incrociata del canale per la preelaborazione e hanno composto la rete di CNN spaziali e temporali collegati in cascata con un codificatore automatico profondo per eseguire la classificazione binaria delle categorie fonologiche. L'accuratezza media ottenuta è stata dell'83,42% nelle sei diverse classificazioni fonologiche binarie. In Cooney et al. [21], hanno utilizzato l'analisi delle componenti indipendenti (ICA) con precondizionamento dell'approssimazione dell'Assia e CNN per classificare 5 vocali immaginarie ed è stata ottenuta un'accuratezza del 32,35%. Nel loro recente lavoro utilizzando ICA con precondizionamento di approssimazione hessiana, DeepConvNet [25] e ShallowConvNet [25] sono stati applicati [26].

Hanno raggiunto una precisione del 62,37% per la classificazione delle coppie di parole. Sebbene siano stati applicati vari metodi di deep learning, non sono ancora stati in grado di funzionare bene come nel paradigma BMI endogeno come MI.

Di solito, DeepConvNet [25] e ShallowConvNet [25] sono comunemente usati per l'analisi del segnale EEG grezzo. Tuttavia, in Cooney et al. [26], hanno utilizzato metodi aggiuntivi di pre-elaborazione o estrazione delle caratteristiche. Pochi studi utilizzano l'apprendimento end-to-end per il discorso immaginario [27]. La pre-elaborazione o l'estrazione di funzionalità può portare alla perdita di informazioni. Queste procedure richiedono anche un costo computazionale. Pertanto, il metodo di apprendimento end-to-end con minima preelaborazione o estrazione delle caratteristiche sta aumentando l'interesse per la classificazione EEG. Inoltre, uno dei problemi importanti nella classificazione EEG è la mancanza di prove e l'alto

dimensionalità [28]. Pertanto, trovare un metodo di apprendimento profondo adatto per la classificazione EEG utilizzando segnali EEG grezzi sta emergendo anche come un nuovo interesse nel campo del BMI.

Per risolvere i suddetti problemi, abbiamo utilizzato il codificatore di rete neurale siamese. La rete neurale siamese prende due input e addestra la stessa rete in parallelo [28]. Ha lo scopo di ottimizzare la distanza nello spazio di incasso. Ad esempio, se due campioni selezionati a caso appartengono alla stessa classe, si apprende che si trovano vicini l'uno all'altro. E se sono classi diverse, imparano a trovarsi lontane l'una dall'altra. La distanza euclidea viene utilizzata per misurare la distanza degli embedding estratti dalla rete neurale siamese [28].

In questo studio, abbiamo proposto un metodo di classificazione basato su codificatore di rete neurale siamese end-to-end. Per quanto ne sappiamo, non esiste un approccio per classificare il discorso immaginario utilizzando il codificatore della rete neurale siamese. Il set di dati è composto da 6 parole, basate sui segnali EEG per il discorso immaginato. Gli incorporamenti vengono estratti in base alla rete neurale siamese e k-nearest neighbor (k-NN) è stato utilizzato per la classificazione. Questo approccio può estrarre le caratteristiche discriminanti corrispondenti a ciascuna classe dai dati grezzi. Pertanto, il metodo proposto può aumentare le prestazioni del discorso immaginato per le sue applicazioni in ambienti di vita reale.

II. MATERIALI E METODI

A. Descrizione dei dati

Il set di dati utilizzato per questo studio è l'open data di Coretto et al [19]. Consiste dei segnali EEG di 15 soggetti. I segnali EEG sono stati registrati utilizzando elettrodi Ag-AgCl. Gli elettrodi sono stati posizionati su F3, F4, C3, C4, P3 e P4 mentre gli elettrodi di riferimento e di terra sono stati posizionati rispettivamente sui mastoidi sinistro e destro. Questi elettrodi includevano parzialmente l'area di Wernicke relativa all'elaborazione del linguaggio [15, 22]. I soggetti hanno eseguito il discorso immaginato e il discorso pronunciato di 5 vocali spagnole e 6 parole spagnole. In questo studio, utilizziamo solo segnali EEG vocali immaginari corrispondenti a 6 parole. Il numero di prove vocali immaginate eseguite da ciascun soggetto era di 40 per parola. Le parole di comando erano arriba (su), abajo (giù), derecha (destra), izquierda (sinistra), adelante (avanti) e atrás (indietro). Questi comandi sono stati scelti per controllare in modo intuitivo i dispositivi esterni in un sistema BMI.

Il paradigma dell'esperimento consiste in 4 procedure per prova come mostrato in Fig. 1. In ogni prova, viene presentato prima un intervallo pronto per 2 sec. In questa procedura, i soggetti sono istruiti a rilassarsi. Quindi vengono presentati segnali visivi e uditivi per 2 sec. Il segnale visivo viene presentato sul monitor e il segnale acustico viene fornito attraverso gli altoparlanti. Quando il segnale scompare, ai soggetti viene chiesto di immaginare la pronuncia della parola presentata

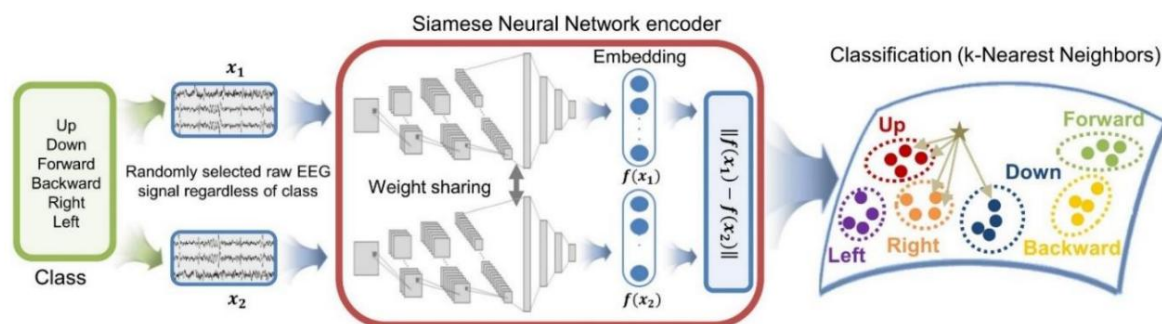


Fig. 2. Quadro proposto di classificazione basata su codificatore di rete neurale siamese end-to-end. Una prova nel set di dati di test è contrassegnata da una stella. Quando la prova di prova si trova nello spazio di incorporamento tramite la rete neurale siamese, la classe viene determinata in base alla classe delle cinque istanze più vicine.

TABELLA I. ARCHITETTURA NEURALE PER LA RETE NEURALE SIAMESE

Strato	Ingresso	Produzione	Nocciolo
Conv1	6 × 512 × 1	2 × 103 × 64	5 × 5
Maxpool1	2 × 103 × 64	1×52×64	2 × 2
Conv2	1×52×64	1 × 52 × 128	1 × 3
Maxpool2	1 × 52 × 128	1 × 52 × 128	1 × 3
Conv3	1 × 52 × 128	1 × 52 × 128	1 × 2
Maxpool3	1 × 52 × 128	1 × 52 × 128	1 × 2
Appiattare	1 × 52 × 128	6656	-
FC1	6656	1024	-
FC2	1024	512	-
FC3	512	256	-
FC4	256	8	-

dalla stecca per 4 sec. La procedura finale è un periodo di riposo per 4 sec. Queste quattro procedure compongono una prova e questa prova viene ripetuta 40 volte per ogni parola in modo casuale.

I segnali EEG sono campionati a 1.024 Hz, questo segnale ad alta dimensione è sottocampionato a 128 Hz. Il down-sampling mira a ridurre la quantità di dati [19]. Il set di dati originale era filtrato passa-banda tra 2 e 40 Hz utilizzando un filtro passa-banda con risposta all'impulso finito (FIR). In questo studio non è stata eseguita alcuna ulteriore elaborazione, come la preelaborazione o l'estrazione di caratteristiche.

B. Metodo proposto

1) *Quadro generale:* il diagramma di flusso generale è mostrato in Fig. 2. In primo luogo, i dati grezzi vengono immessi nella rete neurale siamese. Ogni dato consiste di 4sec e 6 canali. È necessario utilizzare due input sulla rete neurale siamese. Pertanto, vengono selezionati due campioni casuali dai dati totali. Questi due campioni vengono selezionati indipendentemente dalla classe. Ogni campione addestra un ramo CNN, entrambi i rami hanno la stessa struttura e gli stessi parametri. Se le classi dei due dati sono le stesse, si apprende che l'incorporamento estratto attraverso la rete neurale siamese è vicino l'uno all'altro, e nel caso in cui le classi dei due dati sono diverse, si apprende che l'incorporamento è distante.

Di conseguenza, la rete neurale siamese apprende le distanze tra i campioni nello spazio specifico. Utilizzando reti neurali siamesi su dati grezzi ad alta dimensione, possiamo ottenere incorporamenti dimensionalmente ridotti che contengono informazioni importanti relative alla distanza tra classi diverse.

Per la classificazione è stato utilizzato k-NN; prevede le classi di set di dati di test attraverso le k istanze più vicine in uno spazio specifico.

2) *Estrazione delle caratteristiche utilizzando la rete neurale siamese:* abbiamo utilizzato la rete neurale siamese per affrontare l'inconveniente di una piccola quantità di prove EEG. A causa del numero limitato di prove sui dati EEG, è più probabile che l'utilizzo di metodi convenzionali di deep learning causi overfitting. D'altra parte, la rete neurale siamese appartiene all'apprendimento metrico. La rete neurale siamese apprende da due input ed estrae l'incorporamento ridotto. Le due reti sono identiche e condividono i parametri. Gli input sono selezionati casualmente nel training set.

Questi ingressi selezionati casualmente sono indicati con 1, 2. Gli incorporamenti estratti attraverso la rete sono anche indicati come (1),

(2). La rete è addestrata a distinguere se i due input sono uguali o meno [18]. Pertanto, y è un'etichetta che indica se l'input è nella stessa classe (= 1) o in una classe diversa (= 0). Per addestrare la rete neurale siamese, le informazioni relative alle etichette devono essere fornite alla rete in ogni iterazione. La struttura informativa è [] che indica i due input e la corrispondente label. Quando l'incorporamento viene estratto attraverso la rete, la distanza euclidea tra questi due incastri viene calcolata come segue:

$$D(1, 2) = \|f(1) - f(2)\| \quad (y)$$

Molti metodi di deep learning utilizzano la perdita di entropia incrociata. Tuttavia, le perdite di entropia incrociata si concentrano sulla previsione della classe a cui appartiene il campione. Pertanto, non è adatto per l'apprendimento della distanza tra l'incorporamento [29]. La rete neurale siamese utilizza la perdita di contrasto. La perdita contrastiva apprende che i campioni appartenenti alla stessa classe sono posizionati uno vicino all'altro e i campioni appartenenti a classi diverse sono posizionati lontano nello spazio di inclusione [18, 28]. La funzione di perdita contrastiva è la seguente:

$$L(1, 2, y) = \frac{1}{2} \|f(1) - f(2)\|^2 + \frac{1}{2} \max(0, m - \|f(1) - f(2)\|) \quad (y)$$

dove m denota un margine ($m > 0$). Questo è un parametro impostato dall'utente e svolge un ruolo nel rendere la distanza tra due campioni maggiore del suo valore quando due campioni appartengono a classi diverse [29]. La rete neurale siamese era originariamente

proposto per l'apprendimento one-shot [30], al contrario, questo studio mira ad aumentare le prestazioni imparando quante più combinazioni possibili di input nel training set. Utilizziamo l'incorporamento ottenuto dalla rete neurale siamese per addestrare il classificatore; il nostro scopo è classificare i campioni in classi diverse, non determinare se due classi sono uguali o diverse.

La nostra architettura di rete neurale siamese è specificata nella Tabella I. Lo schema del codificatore è stato addestrato utilizzando la rete neurale siamese. La rete neurale siamese è stata addestrata utilizzando l'ottimizzatore ADAM e la funzione di perdita contrastiva. Impostiamo il fattore di decadimento del gradiente su 0,9 e il fattore di decadimento del gradiente al quadrato su 0,99. Impostiamo anche il margine $m=0.5$, il tasso di apprendimento è 0.001 e il numero di epoche è 100 . La funzione di attivazione rettificata (ReLU) è stata utilizzata dopo ogni strato di convoluzione e strato completamente connesso.

3) *Classificazione utilizzando k-NN*: dalla rete neurale siamese, otteniamo un'incorporamento di dimensioni ridotte dai dati originali ad alta dimensione. Il k-NN viene utilizzato per prevedere le classi utilizzando l'incorporamento estratto [31]. Il numero più vicino di vicini è impostato su 5 nel metodo k-NN ($k = 5$). k-NN viene utilizzato perché effettua previsioni basate sull'istanza k più vicina nello spazio di incorporamento specifico. La rete neurale siamese apprende anche la distanza tra i dati in specifici spazi incorporati, quindi k-NN è più adatto per l'analisi rispetto ad altri metodi di classificazione.

C. Schema di formazione e valutazione Per

valutare le prestazioni della classificazione, è stata eseguita una convalida incrociata di 5 volte per ciascun soggetto. In ogni classe, lo stesso numero di prove viene utilizzato per l'allenamento. Altri metodi vengono valutati con lo stesso approccio di validazione incrociata 5 volte per ciascun soggetto. Come misure di classificazione, abbiamo calcolato la precisione, il richiamo, il punteggio F1 e l'accuratezza per la classe 6.

D. Analisi statistica

Abbiamo effettuato un'analisi statistica per stimare la differenza significativa tra il nostro metodo proposto e i diversi metodi. È stata utilizzata l'analisi della varianza unidirezionale (ANOVA) e l'analisi post-hoc è stata eseguita utilizzando il test t appaiato. Per confronti multipli è stata applicata la correzione di Bonferroni. Tutti i valori significativi sono stati fissati a $p\text{-value} = 0,05$.

III. RISULTATI

A. Prestazioni di classificazione utilizzando il metodo proposto

La tabella II mostra l'accuratezza della classificazione utilizzando il metodo proposto in tutte le materie. La classificazione in 6 classi che utilizza il metodo proposto mostra un'accuratezza media del $31,40 \pm 2,73\%$ in tutte le materie. Il soggetto 4 ha mostrato le prestazioni più elevate ($36,51 \pm 4,13\%$).

Fig. 3. illustra la matrice di confusione. Si osserva che "up" mostra le migliori prestazioni di classificazione. Tuttavia, tutte le parole hanno prestazioni simili. Inoltre, tutte le accuratèzze ottenute erano superiori al livello casuale (16,67%). La tabella III mostra il ricordo e il punteggio F1 in tutti i soggetti. La classe "avanti" ha mostrato le prestazioni più elevate, mentre "su" ha mostrato le prestazioni più basse sia nel richiamo che nel punteggio F1.

TABELLA II. PRECISIONE DI CLASSIFICAZIONE DI 6 CLASSI UTILIZZANDO PROPOSTA METODO IN TUTTE LE MATERIE

Soggetto	Precisione (%)
Sub01	$28,84 \pm 2,51$
Sub02	$34,43 \pm 3,15$
Sub03	$30,77 \pm 1,90$
Sub04	$36,51 \pm 4,13$
Sub05	$30,95 \pm 1,06$
Sub06	$33,93 \pm 3,66$
Sub07	$32,69 \pm 2,51$
Sub08	$31,64 \pm 2,11$
Sub09	$27,14 \pm 1,20$
Sub10	$33,82 \pm 2,23$
Sub11	$29,49 \pm 1,67$
Sub12	$29,16 \pm 0,76$
Sub13	$29,69 \pm 1,71$
Sub14	$27,38 \pm 0,65$
Sub15	$34,55 \pm 3,04$
Media \pm Std.	$31,40 \pm 2,73$

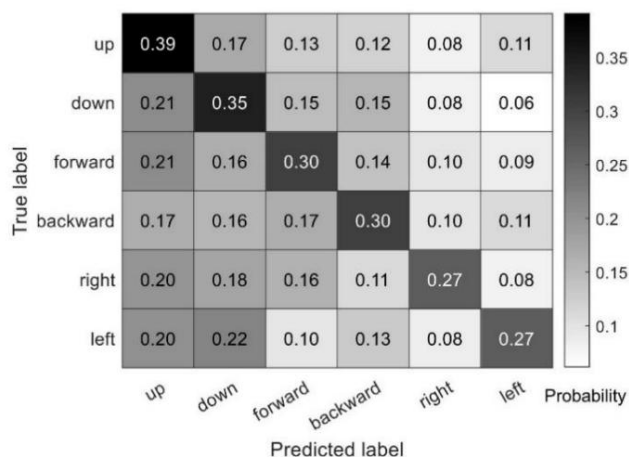


Fig. 3. Matrice di confusione media utilizzando il metodo proposto per tutte le materie.

B. Confronto di metodi convenzionali

Abbiamo confrontato le prestazioni di classificazione del quadro proposto e metodi convenzionali. Il nostro metodo ha migliorato le prestazioni dal 9,93 al 13,94% rispetto all'altro metodo di classificazione di base. La tabella IV mostra l'accuratezza media della classificazione dei diversi metodi in tutte le materie e i risultati dell'analisi statistica. C'erano differenze significative tra le precisioni dei diversi metodi ($F(4) = 164,43$, $p\text{-value} < 0,001$). Inoltre, il test post-hoc ha rivelato differenze significative tra il metodo proposto e altri quattro metodi convenzionali ($p\text{-value} < 0,001$).

IV. DISCUSSIONE

Il metodo di classificazione basato sul codificatore della rete neurale siamese end-to-end proposto mostra che la rete neurale siamese può apprendere caratteristiche discriminanti dai dati grezzi. In particolare, mostra prestazioni migliori rispetto ai metodi di apprendimento automatico che utilizzano la preelaborazione o l'estrazione di funzionalità. I metodi tradizionali di machine learning hanno molte limitazioni quando si addestrano i modelli utilizzando dati grezzi. Pertanto, la pre-elaborazione o feature

TABELLA III. RICHIAMO E MISURA F1 PER OGNI PAROLA DEL DISCORSO IMMAGINATA (CLASSE 6) IN TUTTE LE MATERIE

Soggetto	Su		Giù		Giusto		Sinistra		Inoltre		Indietro	
	Richiamare	Richiamo del punteggio F1	Punteggio F1	Richiamare	Punteggio F1	Richiamare	Punteggio F1	Richiamare	Punteggio F1	Richiamare	Punteggio F1	
Sub01	0,368	0,389	0,250	0,200	0,222	0,229	0,316	0,333	0,278	0,270	0,278	0,294
Sub02	0,214	0,250	0,333	0,400	0,571	0,444	0,333	0,316	0,500	0,375	0,300	0,300
Sub03	0,278	0,345	0,214	0,222	0,294	0,345	0,333	0,320	0,429	0,286	0,400	0,320
Sub04	0,308	0,364	0,500	0,444	0,333	0,364	0,333	0,300	0,500	0,421	0,308	0,320
Sub05	0,154	0,200	0,250	0,167	0,286	0,308	0,300	0,353	0,750	0,545	0,500	0,364
Sub06	0,235	0,348	0,250	0,286	0,308	0,348	0,600	0,429	0,400	0,286	0,750	0,353
Sub07	0,214	0,260	0,267	0,381	0,333	0,267	0,500	0,400	0,429	0,375	0,500	0,357
Sub08	0,231	0,250	0,250	0,333	0,364	0,348	0,273	0,231	0,357	0,370	0,500	0,357
Sub09	0,333	0,300	0,417	0,417	0,333	0,286	0,143	0,188	0,200	0,182	0,333	0,286
Sub10	0,333	0,381	0,250	0,261	0,417	0,417	0,235	0,276	0,500	0,400	0,429	0,316
Sub11	0,278	0,345	0,194	0,293	0,182	0,174	0,600	0,316	0,333	0,182	0,714	0,455
Sub12	0,211	0,286	0,333	0,370	0,125	0,105	0,286	0,200	0,455	0,417	0,333	0,308
Sub13	0,333	0,364	0,250	0,273	0,214	0,250	0,333	0,381	0,429	0,286	0,286	0,222
Sub14	0,211	0,242	0,236	0,258	0,294	0,345	0,267	0,276	0,364	0,308	0,400	0,200
Sub15	0,190	0,286	0,182	0,200	0,250	0,250	0,714	0,588	1,000	0,286	0,667	0,533
Media	0,259	0,307	0,278	0,300	0,302	0,298	0,371	0,327	0,461	0,333	0,446	0,328
Std.	0,062	0,057	0,082	0,084	0,102	0,087	0,153	0,096	0,187	0,093	0,152	0,079

TABELLA IV. RISULTATI STATISTICI IN CLASSIFICAZIONE MEDIA PRESTAZIONI RISPETTO AL METODO PROPOSTO

Metodo	Precisione (%)	valore t	valore p
Coretti et al. [17]	17,46 ± 0,71	-18.81	<0,001
Coney et al. [19]	18,89 ± 1,41	-15.08	<0,001
Schirmeister et al. [23]	19,81 ± 2,10	-14.42	<0,001
Garcia-Salinas et al. [18]	21,47 ± 1,99	-14.43	<0,001
Metodo proposto	31,40 ± 2,73	-	-

il processo di estrazione è assolutamente necessario prima della classificazione o del raggruppamento [29]. Tuttavia, non è facile trovare un metodo di pre-elaborazione o di estrazione delle caratteristiche adatto. È dovuto al fatto che quando il tipo di dati è diverso, cambiano anche le caratteristiche corrispondenti. Il metodo proposto non richiede una procedura di pre-elaborazione o di estrazione delle caratteristiche poiché la rete neurale siamese funziona come un estrattore di caratteristiche. Questo perché il deep learning può apprendere direttamente dati grezzi ad alta dimensione [29].

In particolare, il nostro metodo proposto ha superato i metodi CNN di base. La CNN tradizionale utilizza la perdita di entropia incrociata, che si limita ad apprendere la distanza tra la probabilità reale e quella prevista. D'altra parte, la perdita di contrasto calcola la distanza tra i dati di incorporamento, impara a incorporare campioni vicini l'uno all'altro della stessa classe e ad essere incorporati a una distanza maggiore dei campioni marginali di classi diverse [28-31].

Inoltre, il limite degli studi EEG è che c'è un piccolo numero di dati. Se il numero di dati è troppo piccolo, un approccio che utilizza il deep learning non è appropriato. Tuttavia, è difficile ottenere una grande quantità di dati quando si utilizzano i segnali EEG a causa della natura degli esperimenti, specialmente nei paradigmi del linguaggio immaginato. Ci sono solo 40 prove per classe nel database utilizzato [19]. In alcuni studi sono state registrate solo 12 prove per classe [32]. In questo caso, il metodo di apprendimento profondo con entropia incrociata non apprende bene e tende a sovradimensionarsi. Tuttavia, poiché la rete neurale siamese utilizzava originariamente la perdita di contrasto

proposto in one-shot learning, può essere utilizzato anche per una piccola quantità di dati. La rete neurale siamese utilizza campioni di due coppie come input. Pertanto, ha il vantaggio di poter aumentare il numero di dati del numero di combinazioni [30]. Lo studio ha utilizzato quante più combinazioni possibili per aumentare le prestazioni.

Il livello è inoltre costituito da tre livelli di convoluzione, tre livelli di max-pooling e quattro livelli completamente connessi. Il motivo per utilizzare quattro livelli completamente connessi è che agiscono come una rete neurale profonda, che apprende caratteristiche più discriminanti. La nostra rete ha funzionato meglio rispetto a quando abbiamo applicato la perdita di contrasto a EEGNet [1] o DeepConvNet [25], che sono spesso utilizzati nell'analisi EEG. Di conseguenza, la nostra rete è più adatta per l'analisi del discorso immaginato.

Di conseguenza, il nostro metodo ha mostrato un miglioramento di oltre il 9,93% rispetto ai metodi di riferimento. Inoltre, l'analisi statistica ha mostrato differenze significative tra i nostri metodi proposti e altri. La performance media tra i soggetti è stata del 31,40 ± 2,73%. Il soggetto 4 ha mostrato la performance più alta con 36,51 ± 4,13%. Pertanto, confermiamo che la rete neurale siamese riduce i dati ad alta dimensione (in particolare a 8 dimensioni) riducendo al minimo la perdita di informazioni. Inoltre, la nostra rete neurale siamese classifica i segnali EEG dal parlato immaginario senza rimuovere il rumore extra (ad esempio, la rimozione dell'elettrooculografia). I segnali EEG hanno un basso rapporto segnale/rumore [28], che non può essere risolto con la sola preelaborazione.

Tuttavia, il nostro metodo sembra in parte superarlo.

V. CONCLUSIONE

In questo studio, abbiamo studiato l'approccio di classificazione basato sull'encoder della rete neurale siamese end-to-end per classificare il discorso immaginario usando l'EEG. I dati consistevano in EEG per il discorso immaginario di 6 classi (parola). Il nostro approccio consisteva nell'addestrare la rete neurale siamese utilizzando dati grezzi e classificare i dati di incorporamento ottenuti.

Molti studi precedenti miravano a migliorare le prestazioni del discorso immaginario, tuttavia, la maggior parte di essi si concentra sulla preelaborazione o sui metodi di estrazione delle caratteristiche. Inoltre, tra gli studi sul deep learning, sono pochi gli approcci che utilizzano i dati grezzi come input per la rete. Proponiamo un approccio per migliorare le prestazioni di classificazione utilizzando dati grezzi. Inoltre, abbiamo studiato un approccio di deep learning che può apprendere bene senza overfitting anche quando si utilizza un numero limitato di prove.

In conclusione, abbiamo ottenuto un'accuratezza media del $31,40 \pm 2,73\%$ in più del $9,93\%$ in più rispetto ai metodi di riferimento. I risultati hanno mostrato che la rete neurale siamese basata sulla perdita di contrasto ha il potenziale per classificare il discorso immaginario. La rete neurale siamese è un approccio basato sull'apprendimento metrico profondo. Abbiamo deciso di esaminare altri metodi di apprendimento delle metriche che riteniamo abbiano il potenziale per aumentare le prestazioni. Inoltre, la rete neurale siamese funge da metodo di estrazione delle caratteristiche e k-NN impara a classificare i vettori di incorporamento ottenuti tramite la rete neurale siamese. Il nostro metodo ha mostrato la possibilità di utilizzare il paradigma del discorso immaginato come un BMI intuitivo per ambienti di vita reale.

RICONOSCIMENTO

Gli autori ringraziano J. Kalafatovich e D.-K. Han per l'utile discussione.

RIFERIMENTI

[1] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung e B.J. Lance, "EEGNet: una rete neurale convoluzionale compatta per interfacce cervello-computer basate su EEG", *J. Neural Eng.*, vol. 15, n. 5, giugno 2018.

[2] M. Lee, R.D. Sanders, S.-Y.K. Yeom, D.-O. Vinto, K.-S. Seo, H.-J. Kim, G. Tononi, e S.-Y.W. Lee, "Proprietà della rete nelle transizioni di coscienza durante la sedazione indotta da propofol", *Sci. Rep.*, vol. 7, n. 1, pp. 1-13, dicembre 2017.

[3] X. Ding e S.-W. Lee, "Cambiamenti della connettività funzionale ed efficace nel rifornimento di fumo su fumatori accaniti privati: uno studio fMRI in stato di riposo", *PLoS One*, vol. 8, n. 3, marzo 2013.

[4] D.-O. Vinto, H.-J. Hwang, D.-M. Kim, K.-Y.R. Müller e S.-Y.W. Lee, "Presentazione visiva seriale rapida basata sul movimento per interfacce cervello-computer indipendenti dallo sguardo", *IEEE Trans. Sistema neurale Riabilitazione. Ing.*, vol. 26, n. 2, pp. 334-343, agosto 2018.

[5] Y. Chen, A.D. Atnafu, I. Schlattner, W.T. Weltdsadiq, M.-C. Roh, H.J. Kim, S.W. Lee, B. Blankertz e S. Fazli, "Un sistema di accesso basato su EEG ad alta sicurezza con stimoli RSVP ed elettrodi asciutti", *IEEE T. Inf. Sicurezza forense*, vol. 11, n. 12, pp. 2635-2647, giugno 2016.

[6] L. Wang, X. Liu, Z. Liang, Z. Yang e X. Hu, "Analisi e classificazione del BCI ibrido basato su immagini motorie e immagini vocali", *Misura*, vol. 147, luglio 2019.

[7] L.F. Nicolas-Alonso e J. Gomez-Gil, "Interfacce cervello-computer, una recensione", *Sensors*, vol. 12, n. 2, pp. 1211-1279, gennaio 2012.

[8] K.-T. Kim, H.-I. Suk e S.-Y.W. Lee, "Comando di una sedia a rotelle controllata dal cervello utilizzando potenziali evocati somatosensoriali allo stato stazionario", *IEEE Trans. Sistema neurale Riabilitazione. Ing.*, vol. 26, n. 3, pp. 654-665, agosto 2016.

[9] M.-H. Lee, J. Williamson, D.-O. Vinto, S. Fazli, e S.-W. Lee, "Un sistema di ortografia ad alte prestazioni basato su segnali EEG-EOG con feedback visivo", *IEEE Trans. Sistema neurale Riabilitazione. Ing.*, vol. 26, n. 7, pagg. 1443-1459, maggio 2018.

[10] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli e S.-Y.W. Lee, "Set di dati EEG e toolbox OpenBMI per tre paradigmi BCI: un'indagine sull'analfabetismo BCI", *GigaScience*, vol. 8, n. 5, maggio 2019.

[11] H.-I. Suk e S.-W. Lee, "Selezione di bande di frequenza specifiche per soggetto e classe per la classificazione di immagini motorie multiclasse", *Int. J. Imaging Syst. Technol.*, vol. 21, n. 2, pp. 123-130, maggio 2011.

[12] S.-H. Lee, M. Lee e S.-Y.W. Lee, "Rappresentazioni EEG di caratteristiche spaziali e temporali nel discorso immaginario e nel discorso palese", *In Asian Conference on Pattern Recognition*, 2019, novembre pp. 387-400.

[13] S. Perdikis, R. Leeb, J. Williamson, A. Ramsay, M. Tavella, L. Desideri e J.D.R. Millán, "Clinical evaluation of braintree, a motor imagery hybrid BCI speller", *J. Neural Eng.*, vol. 11, n. 3, pp. 036003, aprile 2014.

[14] S.-H. Lee, M. Lee, J.-H. Jeong e S.-Y.W. Lee, "Verso un sistema di comunicazione BCI intuitivo basato su EEG che utilizza parole immaginate e immagini visive", *In IEEE International Conference on Systems, Man and Cybernetics (SMC)*, ottobre 2019, pp. 4409-4414.

[15] C.H. Nguyen, G.K. Karavas e P. Artemiadis, "Dedurre il discorso immaginario utilizzando i segnali EEG: un nuovo approccio utilizzando le molteplici caratteristiche riemanniane", *J. Neural Eng.*, vol. 15, n. 1, pp. 016002-016018, febbraio 2018.

[16] C. Cooney, R. Folli e D. Coyle, "Ricerca neurolinguistica che promuove lo sviluppo di un'interfaccia cervello-computer a discorso diretto", *iScience*, vol. 8, pp. 103-125, ottobre 2018.

[17] C.S. Dasalla, H. Kambara, M. Sato e Y. Koike, "Classificazione a prova singola delle immagini del parlato vocale utilizzando modelli spaziali comuni", *Rete neurale*, vol. 22, n. 9, pp. 1334-1339, novembre 2009.

[18] S. Shahtalebi, A. Asif e A. Mohammadi, "Siamese neural networks for EEG-based brain-computer interfaces", *arXiv preprint arXiv2002.00904*, 2020.

[19] G.A.P. Coretto, I.E. Gareis e H.L. Rufiner, "Database ad accesso aperto di segnali EEG registrati durante il discorso immaginario", *International Symposium on Medical Information Processing and Analysis*, vol. 10160, gennaio 2017, pp. 1016002.

[20] J.S. García-Salinas, L. Villaseñor-Pineda, C.A. Reyes-García, e A.A. Torres-García, "Trasferisci l'apprendimento nei BCI basati su EEG del linguaggio immaginario", *Biomed. Processo di segnale. Controllo*, vol. 50, pp. 151-157, aprile 2019.

[21] C. Cooney, R. Folli e D. Coyle, "L'ottimizzazione dei livelli migliora la generalizzazione della CNN e trasferisce l'apprendimento per la decodifica del parlato immaginato dall'EEG", *In IEEE International Conference on Systems, Man and Cybernetics (SMC)*, ottobre 2019, pp. 1311-1316.

[22] M.N.I. Qureshi, B. Min, H.J. Park, D. Cho, W. Choi e B. Lee, "Classificazione multiclasse del discorso dell'immaginazione delle parole con funzionalità di connettività ibrida", *IEEE Trans. Biomed. Ing.*, vol. 65, n. 10, pagg. 2168-2177, ottobre 2017.

[23] P. Saha, S. Fels e M. Abdul-Mageed, "Apprendimento approfondito del collettore EEG per la categorizzazione fonologica dai pensieri attivi", *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, maggio 2019, pp. 2762-2766.

[24] P. Saha, M. Abdul-Mageed e S. Fels, "Di quello che pensi! verso il riconoscimento vocale immaginato con il deep learning gerarchico", *arXiv preprint arXiv1904.05746*, 2019.

[25] R.T. Schirrmester, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggenesperger, M. Tangermann e T. Ball, "Deep learning con reti neurali convoluzionali per la decodifica e la visualizzazione EEG", *Hum. Mappa del cervello*, vol. 38, n. 11, pp. 5391-5420, agosto 2017.

[26] C. Cooney, A. Korik, F. Raffaella e D. Coyle, "Classificazione di coppie di parole parlate immaginate utilizzando reti neurali convoluzionali", *Conferenza BCI di Graz*, settembre 2019, pp. 338-343.

[27] P. Sun e J. Qin, "Neural networks based EEG-speech models", *arXiv preprint arXiv1612.05369*, 2016.

[28] S. Stober, "Apprendimento delle caratteristiche discriminative dalle registrazioni elettroencefalografiche codificando i vincoli di somiglianza", *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, marzo 2017, pp. 6175-6179.

[29] M. Kaya e H.S. Bilge, "Deep metric learning: a survey", *Symmetry*, vol. 11, n. 9, pp. 1066, agosto 2019.

[30] G. Koch, R. Zemel e R. Salakhutdinov, "Reti neurali siamesi per il riconoscimento di immagini one-shot", *In ICML Deep Learning Workshop*, vol. 2, luglio 2015.

[31] R. Thiyagarajan, C. Curro e S. Keene, "Uno spazio di incorporamento appreso per il clustering del segnale EEG", *in IEEE Signal Processing in Medicine and Biology Symposium*, dicembre 2017, pp. 1-4.

[32] S. Zhao e F. Rudzicz, "Classificazione delle categorie fonologiche nel discorso immaginato e articolato", *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, aprile 2015, pp. 992-996.