

تشخیص تصور گفتار کلمات بازی سنگ، کاغذ، قیچی با استفاده از سیگنال‌های EEG

مجید مقدری^۱، دانشجوی دکترا، مینا زلفی لبقوان^۲، استادیار، سبلان دانشور^۳، استادیار

۱- دانشکده مهندسی برق و کامپیوتر - دانشگاه تبریز - تبریز - ایران - m.moghaddari@tabrizu.ac.ir

۲- دانشکده مهندسی برق و کامپیوتر - دانشگاه تبریز - تبریز - ایران - mzolffy@tabrizu.ac.ir

۳- دانشکده مهندسی برق و کامپیوتر - دانشگاه برونل - لندن - انگلستان - sebelan.danishvar@brunel.ac.uk

چکیده: درک صحبت‌های افراد بدون نیاز به بیان و تنها با استفاده از رمزگشایی سیگنال‌های مغزی در هنگام تصور گفتار، یکی از بروزترین تحقیقات در زمینه هوش مصنوعی می‌باشد. سومین مسابقه ملی واسط مغز و رایانه که توسط مرکز ملی نقشه برداری مغز ایران در سال ۱۳۹۹ برگزار گردید به دسته‌بندی تصور گفتار برای سه کلمه سنگ، کاغذ و قیچی اختصاص یافت. در این مسابقه نویسندگان این مقاله با استفاده از تجزیه بسته‌های موجک و الگوی مکانی مشترک و بکارگیری دسته‌بندهای مختلف توانستند مقام دوم را کسب نمایند. دسته‌بندهای ماشین بردار پشتیبان، k-نزدیک‌ترین همسایه، جنگل تصادفی، رگرسیون لجستیک، XGBoost و مدل یادگیری عمیق Dense را بصورت جداگانه برای هر فرد و همچنین بصورت همزمان برای تمامی افراد ارزیابی نمودیم و بهترین دقت میانگین ۵۱/۷٪ حاصل شد. در ادامه این مدل را با استفاده از روش اسپکتروگرام و شبکه عصبی کانولوشنی توسعه دادیم و توانستیم به دقت میانگین ۷۶/۵٪ دست یابیم. این دقت از بهترین دقت گزارش شده بر روی این مجموعه داده بسیار بهتر بوده است. همچنین عملکرد این مدل در مقایسه با پژوهش‌های اخیر در این زمینه بر روی مجموعه داده‌های مختلف، برتری دارد.

واژه‌های کلیدی: تصور گفتار، صحبت ذهنی، سیگنال‌های مغزی، واسط مغز-رایانه، EEG

Decoding Speech Imagination for Words of Rock-Paper-Scissors Game using EEG Signals

Majid Moghaddari, PhD student¹, Mina Zolfy Lighvan, Assitant professor², Sebelan Danishvar, Assitant professor³

1- Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran, Email: m.moghaddari@tabrizu.ac.ir

2- Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran, Email: mzolffy@tabrizu.ac.ir

3- Department of Electronic and Computer Engineering, Brunel University, London, UK, Email: sebelan.danishvar@brunel.ac.uk

Abstract: Recognition of silent speech only by decoding brain signals is one of the latest research in the field of artificial intelligence. The 3rd Iranian national brain-computer interface competition, which was held by the National Brain Mapping Center of Iran in 2020, was dedicated to the classification of imagined speech for the three words of rock-paper-scissors game. In this contest, the authors introduced an approach based on wavelet packets decomposition and common spatial pattern and could win the second place. We evaluated some of the most famous classifiers including support vector machine, k-nearest neighbor, random forest, logistic regression, XGBoost and dense deep learning model separately for each subject and simultaneously for all subjects. The best average accuracy was 51.7%. Then we developed the model using spectrogram and convolutional neural network and achieved average accuracy of 76.5%. The accuracy was much better than the accuracy reported by other researchers on this dataset. Also, the performance of our model is superior to recent research in this field on different datasets.

Keywords: Speech imagination, Mental speech, Brain signals, Brain-computer interface, EEG

نام نویسنده مسئول: مینا زلفی لبقوان

نشانی نویسنده مسئول: ایران - تبریز - بلوار بهمن - دانشگاه تبریز - دانشکده مهندسی برق و کامپیوتر

۱- مقدمه

زمانی که آندره مورا نویسنده شهیر فرانسوی رمان تخیلی ماشین خواندن افکار را در سال ۱۹۳۷ نوشت شاید مانند صدها رمان تخیلی دیگر کسی تصور نمی‌کرد که داستان ساخت ماشینی که قادر به استخراج ذهن و تصورات انسان‌ها باشد روزی به واقعیت بپیوندد. اما به کمک پیشرفت هوش مصنوعی و توسعه واسطه‌های مغز-رایانه، خواندن و رمزگشایی اطلاعات ذهنی انسان‌ها ممکن شده است [۱]. یکی از پیشرفت‌های اخیر در زمینه توسعه واسطه‌های مغز-رایانه، تلاش برای تشخیص تصور گفتار در ذهن افراد است. عمل صحبت کردن وقتی رخ می‌دهد که کلماتی که فرد به آن فکر کرده است پس از گذشتن از تارهای صوتی حنجره بصورت صدا از دهان فرد خارج شود. کسانی که به علت مشکلات جسمی قادر به تکلم نیستند مجبورند بجای استفاده از صوت، از اشاره‌های صورت و دست استفاده نمایند. وجود سیستمی برای تشخیص صحبت‌های ذهنی توسط این افراد می‌تواند نقش مهمی در ایجاد و بهبود ارتباط آن‌ها با دیگران و همچنین صدور دستورات لازم به دستگاه‌های هوشمند ایفا نماید. همچنین در شرایطی مانند عملیات‌های حساس نظامی که نیازمند سکوت مطلق می‌باشد امکان گفتگوی ذهنی افراد با یکدیگر می‌تواند بسیار سودمند باشد. از آنجایی که دروغ گفتن به تفاوت آنچه انسان به آن می‌اندیشد و آنچه بر زبان می‌آورد مربوط می‌شود با تشخیص صحبت‌های ذهنی افراد در بسیاری از موارد حساس مانند بازجویی‌ها از تروریست‌ها و مجرمان می‌توان به حقیقت ماجرا پی برد.

۲- تحقیقات پیشین

در سال‌های اخیر زمینه تحقیقاتی تشخیص تصور گفتار مورد توجه قرار گرفته است. تجهیزات مختلفی برای ثبت سیگنال‌ها و تصاویر مغزی به هنگام تصور گفتار، و روش‌های مختلفی برای رمزگشایی از آن‌ها بکار گرفته شده‌اند که مهم‌ترین آن‌ها عبارتند از MEG [۲]، EMG [۳]، fNIRS [۴]، EcOG [۵] و MRI [۶]. در مقایسه با تمامی این تجهیزات، دستگاه‌های الکتروانسفالوگرافی (EEG) با توجه به غیرتهاجمی بودن، قیمت کمتر، سهولت استفاده و جابجایی، استفاده بیشتری در ساخت واسطه‌های مغز-رایانه داشته‌اند. ونگ و همکاران [۷] با استفاده از الگوی مکانی مشترک (CSP) و ماشین بردار پشتیبان بر روی تشخیص و دسته‌بندی تصور گفتار برای دو کاراکتر به زبان چینی توسط سیگنال‌های EEG ثبت شده از هشت شرکت کننده تحقیق نمودند. میانگین دقت دسته‌بندی بین دو کلمه ۶۶/۸۶٪ گزارش شد. با توجه به نتایج، CSP برای دسته‌بندی کلمات چینی عملکرد مطلوبی نداشته است. این می‌تواند به دلیل شباهت سیگنال‌های EEG تصور ذهنی کلمات چینی باشد. همچنین CSP بسیار به نویز و آرتیفکت حساس می‌باشد.

در پژوهشی دیگر [۸] در این زمینه سعی شده است با استفاده از روش‌های مختلف استخراج و انتخاب ویژگی‌ها شامل تجزیه موجک،

تنویری آشوب، بُعد فرکتال، انرژی فرکانس و تجزیه مولفه‌های اصلی، تصور گفتار در سیگنال‌های EEG پیوسته تشخیص داده شود. سه مجموعه داده مختلف مورد ارزیابی قرار گرفت که در مجموعه‌های اول و دوم، بیست و هفت شرکت کننده پنج کلمه اسپانیایی مختلف را در ذهن خود بیان می‌نمودند و در سومی بیست شرکت کننده چهار کلمه اسپانیایی را بصورت ذهنی بیان می‌نمودند. همچنین از دسته‌بندی ماشین بردار پشتیبان، رگرسیون لجستیک، k-نزدیک‌ترین همسایه، و جنگل تصادفی برای دسته‌بندی کلمات استفاده نمودند. رگرسیون لجستیک و بعد از آن ماشین بردار پشتیبان بهترین عملکرد نسبی را در بین دسته‌بندی‌ها داشته‌اند. بهترین دقت بدست آمده برای هر یک از سه مجموعه داده به ترتیب ۶۶٪، ۷۰٪، و ۶۵٪ بوده است. با توجه به این که در سه مجموعه داده این پژوهش، از سه پروتکل ثبت مختلف استفاده شده است تفاوت عملکرد در مجموعه داده‌های اول و دوم تنها به پروتکل ثبت مرتبط بوده است. در مجموعه داده سوم نیز در حالی که تعداد دسته‌ها کاهش یافته بود انتظار می‌رفت که دقت افزایش یابد اما کاهش یافته است. این می‌تواند به دلیل تفاوت پروتکل ثبت باشد و این که تصاویر نشان داده شده به فرد سبب می‌شده است که تصور تصویر نیز در هنگام تصور گفتار ایجاد شود و سبب ایجاد اختلال در تشخیص سیگنال‌های ناشی از تصور گفتار با تصور تصویر شده است.

در پژوهش‌های جدید، محققان سعی نمودند از توانایی‌های مدل‌های یادگیری عمیق برای استخراج خودکار ویژگی‌ها و دسته‌بندی تصور گفتار استفاده نمایند. در یکی از این پژوهش‌ها [۹] محققان سعی نمودند عملکرد مدل‌های یادگیری عمیق در تشخیص تصور گفتار را بر روی مجموعه داده‌ای متشکل از پانزده شرکت کننده که تلفظ شش حرف صدادار و شش کلمه اسپانیایی را در ذهن خود تصور می‌نمودند ارزیابی نمایند. این محققان عملکرد مدل‌های ShallowCNN، CNN و EEGNet را در دسته‌بندی کلمات بررسی نمودند. استخراج ویژگی‌های مطلوب نیز بر عهده مدل یادگیری عمیق گزاریده شده بود و تنها بر روی تنظیم پارامترهای مدل بررسی صورت گرفت. دقت میانگین برای این سه مدل به ترتیب برابر با ۲۴،۸۸٪، ۲۴،۷۸٪ و ۲۴،۹۰٪ به دست آمد. این محققان اعلام نمودند مدل‌های یادگیری عمیق عملکرد بهتری از روش‌های یادگیری ماشین متداول داشته است هر چند نتایج به دست آمده برای کاربردهای واقعی چندان مفید نمی‌باشد.

محققان در پژوهشی دیگر [۱۰] با استفاده از ترکیب دو مدل یادگیری عمیق و مدل تطبیق دامنه بدون ناظر سعی نمودند تا تصور گفتار را در افراد مختلف تشخیص دهند. مدل یادگیری عمیق این محققان از ترکیب نرمال‌سازی دسته‌ای تطبیقی و تابع هزینه مبتنی بر اطلاعات پراکندگی، برای ایجاد یک دسته‌بندی تطبیقی بهره برده است. تصور گفتار مورد تشخیص در پژوهش این محققان از پنج کلمه اسپانیایی تصور شده توسط پانزده شرکت کننده و دو کلمه اسپانیایی طولانی تصور شده توسط شش شرکت کننده اسپانیایی زبان تشکیل شده بود. این محققان روشی به نام تطبیق دامنه اصلاحی استاندارد (SRDA) را معرفی

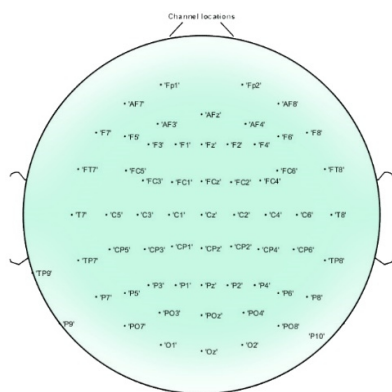
نتایج عملکرد آن‌ها در بخش ۵ نشان داده خواهد شد. در بخش ۶ نیز نتایج حاصل مورد بررسی و تحلیل قرار می‌گیرند و ایده‌هایی برای پژوهش‌های آینده ارائه می‌گردد.

۳- مجموعه داده

ما از مجموعه داده "سومین مسابقه ملی واسط مغز- رایانه" که توسط مرکز ملی نقشه برداری مغز ایران در سال ۱۳۹۹ برگزار گردید استفاده نمودیم. این مجموعه داده بصورت عمومی در دسترس می‌باشد [۱۲].

۳-۱- پروتکل ثبت سیگنال

برای ایجاد این مجموعه داده، سیگنال‌های EEG ۱۵ فرد سالم راست دست (۵ زن و ۱۰ مرد) با میانگین سنی ۳۱ سال ثبت شده است. سیستم ثبت سیگنال دارای ۶۴ کانال با فرکانس نمونه برداری ۲۴۰۰ Hz بوده و فیلتر حذف برق شهر حین ثبت داده فعال بوده است. الکتروده زمین به پیشانی و الکتروده مرجع به گوش راست وصل بوده است (که از مجموعه داده‌ها حذف شده و بنابراین ۶۳ کانال داده وجود دارد). سیگنال همه‌ی الکترودها بین فرکانس‌های ۱ و ۱۳۰ Hz فیلتر شده‌اند. آرایش الکترودها در شکل ۱ نشان داده شده است.



شکل ۱: آرایش الکترودها بر حسب نام آن‌ها

۳-۲- پروتکل آزمایش

فرد بر روی یک صندلی راحت نشسته و صفحه نمایش روبروی او در فاصله‌ی نیم متری قرار دارد. کاری که فرد حین آزمایش باید انجام دهد تصور گفتار است. بدین ترتیب که تصویر دست در حالت‌های مختلف بازی سنگ، کاغذ، قیچی در مانیتور به فرد نشان داده شده و سپس با ظاهر شدن کلمه Go، فرد ۳ ثانیه فرصت دارد تا کلمه متناظر با آن را به صورت بی‌صدا در ذهن خود بیان کند. پروتکل آزمایش در شکل ۲ نشان داده شده است.



شکل ۲: پروتکل آزمایش

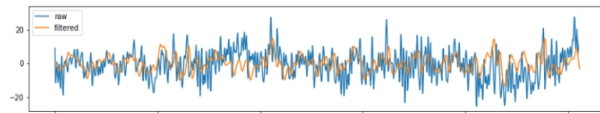
نمودند. دقت‌های به دست آمده در تشخیص کلمات برای دو مجموعه داده ۶۱٪/۶۲٪ و ۹۹٪/۹۶٪ گزارش شده است.

با توجه به پیشرفت‌های صورت گرفته در زمینه تصور گفتار، آزمایشگاه ملی نقشه برداری مغز ایران سومین مسابقه ملی واسط مغز- رایانه (iBCIC2020) را برای اولین بار در کشور ایران به زمینه تشخیص تصور گفتار اختصاص داد. در این مسابقه که پژوهشگرانی از مراکز مختلف دانشگاهی و تحقیقاتی شرکت نموده بودند تشخیص تصور گفتار برای کلمات بازی سنگ، کاغذ، قیچی مورد کنکاش قرار گرفت. با توجه به این‌که تشخیص تصور گفتار برای کلمات فارسی تاکنون بندرت مورد پژوهش قرار گرفته است این مسابقه آغازگر پژوهش جدی و عملیاتی بر روی تشخیص تصور گفتار واژگان فارسی محسوب می‌گردد. ما در این مسابقه توانستیم مقام دوم را بدست آوریم. حامدی و همکاران [۱۱] با استفاده از روش‌های استخراج و انتخاب ویژگی‌های اطلاعات متقابل، الگوی مکانی مشترک و آنالیز مولفه همسایگی، و بکارگیری دسته‌بند یادگیری گروهی توانستند با دقت ۵۱٪/۹۰٪ مقام اول مسابقه را کسب نمایند. ما پس از انجام تغییرات و بهبودهای لازم بر روی روش‌های استخراج و انتخاب ویژگی‌ها و با استفاده از مدل شبکه عصبی کانولوشنی، توانستیم به دقتی بهتر از بهترین دقت بدست آمده در مسابقه بر روی مجموعه داده منتشر شده دست یابیم. ما در این مقاله عملکرد روش‌های استفاده از تجزیه بسته‌های موجک و الگوی مکانی مشترک را که در مسابقه استفاده نموده بودیم و روش اسپکتروگرام را که اکنون توسعه دادیم را در استخراج ویژگی‌های مرتبط مقایسه نمودیم. علاوه بر معرفی بهترین مدل پیشنهادی، برای اولین بار مقایسه‌ای بین عملکرد تعدادی از دسته‌بندهای متداول و مدل‌های یادگیری عمیق در دقت دسته‌بندی تصور گفتار برای کلمات فارسی ارائه نمودیم.

از آنجایی که الگوی سیگنال‌های EEG افراد مختلف برای یک عمل یکسان، می‌تواند بسیار متفاوت باشند لذا مدلی که بر اساس سیگنال‌های ذهنی یک فرد آموزش داده شده و عملکرد مطلوبی دارد ممکن است در دسته‌بندی سیگنال‌های ذهنی فرد دیگر عملکرد خوبی نداشته باشد. به همین منظور در بسیاری از پروژه‌های واسط مغز- رایانه، در ابتدا چندین بار مدل با استفاده از سیگنال‌های فرد جدید آموزش داده می‌شود تا قادر به کشف الگوی اختصاصی این فرد باشد. ما نیز در این پروژه در ابتدا مدل پیشنهادی را برای هر فرد بصورت جداگانه آموزش داده‌ایم و عملکرد مدل را بر روی داده‌های آزمایشی همان فرد ارزیابی نمودیم. سپس در مرحله بعدی، مدل پیشنهادی را بر روی داده‌های آموزشی تمامی افراد بصورت همزمان آموزش دادیم و عملکرد آن را در دسته بندی داده‌های آزمایشی هر فرد و همچنین تمام مجموعه داده ارزیابی نمودیم.

در ادامه مقاله در بخش ۳، توضیحاتی در مورد ویژگی‌های مجموعه داده مورد استفاده و نحوه ثبت آن‌ها ارائه خواهد شد. مراحل پیش‌پردازش و آماده‌سازی داده‌ها در بخش ۴ توضیح داده می‌شود. ساختار و معماری مدل‌های پیشنهادی و دسته‌بندهای مختلف به همراه

این انتخاب به این دلیل بوده است که فرکانس قطع برای فیلتر بالاگذر ۴ هرتز نزدیک به صفر می‌باشد. بنابراین مرتبه‌ای که به فیلتر میان‌گذر باید تخصیص داده می‌شد مناسب برای عبور بالاگذر بوده است اما برای عبور پایین‌گذر نامناسب بوده است. در صورت تلاش برای افزایش مرتبه فیلتر، خروجی فیلتر ناپایدار می‌شده است. در شکل ۴ اثر اعمال فیلتر بر روی سیگنال نشان داده شده است.

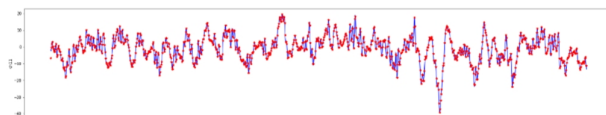


شکل ۴: مقایسه سیگنال بدون فیلتر و فیلتر شده

ب) داده‌های پیوسته با استفاده از ابزار ASR (Artifact Subspace Reconstruction) پاکسازی شد. ابزار ASR کانال‌های الکتریکی دارای داده‌های نویزی، بدون نوسان، دارای نوسان‌های انفجاری کوتاه مدت، و با نوسان‌های فرکانس پایین را از سیگنال EEG پاک می‌نماید. ج) مقادیر جدید به کانال‌های حذف شده در مرحله قبل با استفاده از درونیایی کانال‌های پیرامونی آن‌ها تخصیص داده شد. د) بخاطر کاهش تاثیر خطای ناشی از وجود نویز در الکترومد مرجع در هنگام ثبت داده‌ها، کانال مرجع داده‌ها با میانگین‌گیری از سایر کانال‌ها، به مرجع میانگین تغییر داده شد. همچنین این عمل سبب می‌شود که اطلاعات بی ارزش مشترک در تمام کانال‌ها حذف شود.

۴-۲- کاهش نرخ نمونه‌برداری

نرخ نمونه‌برداری اصلی داده‌ها ۲۴۰۰ هرتز بوده است. ما نرخ نمونه‌برداری را به ۱۰۲۴، ۵۱۲، ۲۵۶ و ۱۲۸ هرتز کاهش دادیم. نتایج ارزیابی مدل با نرخ‌های نمونه‌برداری کاهش‌یافته نشان داد که تغییری در عملکرد مدل در نرخ‌های نمونه‌برداری ۱۰۲۴ و ۵۱۲ هرتز ایجاد نشده است اما کاهش نرخ نمونه‌برداری به ۲۵۶ و ۱۲۸ هرتز با کاهش دقت مدل همراه بوده است. بنابراین در این پژوهش، ما نرخ نمونه‌برداری را به ۵۱۲ هرتز به علت زمان پردازش سریع‌تر و کاهش هزینه‌های محاسباتی با حذف افزونگی‌های داده، کاهش دادیم. شکل ۵ نمونه‌ای از سیگنال با نرخ اصلی و کاهش یافته را نشان می‌دهد.



شکل ۵: نمونه‌ای از سیگنال با نرخ اصلی و کاهش یافته

۴-۳- تفکیک داده‌های آموزشی/آزمایشی

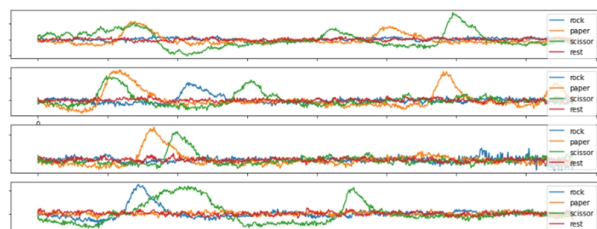
ما از روش از اعتبارسنجی متقابل k -بخشی (k-fold cross validation) برای تفکیک داده‌ها استفاده نمودیم. مجموعه داده در روش اعتبارسنجی متقابل k -بخشی به k بخش متمایز تقسیم می‌شود. $k-1$ بخش برای آموزش مدل و بخش باقیمانده برای آزمایش مدل استفاده می‌گردد. نتیجه نهایی ارزیابی مدل از میانگین نتایج حاصل برای تمام k بخش آزمایشی بدست می‌آید. متداول‌ترین مقدار k در ارزیابی مدل‌ها برابر با

ابتدا علامت + در مرکز صفحه‌ی نمایش به مدت ۲ ثانیه نشان داده می‌شود. در این مدت فرد نباید به موضوعی فکر کند و باید آماده‌ی مشاهده نشان باشد. در ۲ ثانیه بعد، نشانه مورد نظر ظاهر می‌شود. سه نوع نشان در این پروتکل استفاده شده است (متناظر با سنگ، کاغذ و قیچی). پس از گذشت ۲ ثانیه نشان محو می‌شود و کلمه Go به مدت ۰/۵ ثانیه ظاهر و فرد از لحظه‌ی مشاهده آن ۳ ثانیه فرصت خواهد داشت تا تلفظ کلمه متناظر را تصور (بیان بدون صدا در ذهن) کند.

۳-۳- فرمت داده‌ها

داده‌های ۱۵ فرد در اختیار شرکت کنندندهای مسابقه قرار داده شد. داده‌های هر فرد حاوی سیگنال‌های EEG آن فرد از لحظه‌ی ظاهر شدن کلمه "Go" به مدت ۳ ثانیه حین تصور بیان سنگ، ۳ ثانیه حین تصور بیان کاغذ، ۳ ثانیه حین تصور بیان قیچی و ۳ ثانیه در حالت استراحت (فاقد تصور گفتار) است.

در شکل ۳ نمونه‌هایی از سیگنال‌های مجموعه داده برای چهار دسته سنگ، کاغذ، قیچی و استراحت نشان داده شده است.



شکل ۳: نمونه‌هایی از سیگنال‌های مجموعه داده

۴- پیش پردازش و آماده‌سازی داده‌ها

این بخش روش‌های استفاده شده به منظور آماده‌سازی داده‌ها برای دسته‌بندی توضیح داده می‌شود.

۴-۱- پیش پردازش

سیگنال‌های EEG هنگام عبور از بافت سخت جمجمه به شدت تضعیف می‌شوند. نویزهای حاصل از میدان‌های الکترومغناطیسی منابع الکتریکی و آرتیفکت‌های ناشی از پلک زدن، ضربان قلب و حرکات عضلانی به طور قابل توجهی از کیفیت سیگنال‌های EEG می‌کاهند. بطور معمول داده‌ها به منظور حذف نویز و آرتیفکت، با بازبینی توسط یک فرد خبره و یا توسط ابزارهای خودکار پیش‌پردازش می‌شوند. ما در این پژوهش مراحل پیش‌پردازش زیر را بر روی تمامی نمونه‌ها مطابق روش معرفی شده توسط Makoto [۱۳] با استفاده از جعبه ابزار EEGLAB متلب انجام دادیم.

الف) فرکانس‌های زیر ۴ هرتز و بالای ۴۰ هرتز را حذف نمودیم. در ابتدا یک فیلتر FIR بالاگذر ۴ هرتز و سپس یک فیلتر FIR پایین‌گذر ۴۰ هرتز را اعمال نمودیم. بجای اعمال یک فیلتر میان‌گذر ۴-۴۰ هرتز ترجیح دادیم از دو فیلتر پشت سرهم بالاگذر و پایین‌گذر استفاده نماییم.

تجزیه بسته‌های موجک هر دو ضریب تقریب و ضریب جزئی از فیلتر می‌گذرد. برای تجزیه n -سطحی، در حالیکه تبدیل موجک گسسته $n+1$ مجموعه تولید می‌کند WPD به تعداد $2n$ مجموعه مختلف تولید می‌کند. مشکل اصلی DWT در استخراج ویژگی‌های مطلوب این است که پایه‌های موجک در فرکانس‌های پایین با توان دو افزایش می‌یابد در حالیکه مجموعه‌هایی که تولید می‌کند به این نسبت افزایش نمی‌یابد. در WPD تعداد مجموعه‌های تولید شده با ضریب افزایش پایه‌های موجک برابر بوده و در نتیجه نتایج بهتری را حاصل می‌کند.

ما در روش اول، ویژگی‌های داده‌های آموزشی و آزمایشی را توسط WPD پنج سطحی که از موجک های db4 (Daubechies 4 wavelets) استفاده می‌کند استخراج نمودیم. برای هر نمونه داده، ۸ ویژگی تشکیل شده از ۴۶ مجموعه WPD برای هر یک از ۶۳ کانال استخراج شده است.

۲-۱-۵- الگوی مکانی مشترک (CSP)

الگوی مکانی مشترک به عنوان یکی از موفق‌ترین الگوریتم‌های استخراج ویژگی در واسط‌های مغز-رایانه شناخته می‌شود. روش CSP بر مبنای تجزیه سیگنال‌ها به زیرمولفه‌ها به گونه‌ای است که تفاوت آنها را ماکزیمم کند. CSP برای اولین بار به منظور دسته‌بندی تصور حرکت دست در سیگنال‌های EEG مورد استفاده قرار گرفت [۱۵] و پس از آن تبدیل به مهم‌ترین و متداول‌ترین روش انتخاب ویژگی و انتخاب کانال در تحلیل سیگنال‌های EEG شد. CSP فیلترهای مکانی را با حل مسأله‌ی تجزیه‌ی مقادیر ویژه که شامل کواریانس‌های مکانی میانگین برای هر یک از دسته‌ها می‌شود بدست می‌آورد. ما یک CSP با ۴ مولفه مبتنی بر اطلاعات متقابل بر روی هر یک از ۸ ویژگی استخراج شده از WPD در هر نمونه اعمال نمودیم و در نتیجه برای هر نمونه ۳۲ ویژگی نهایی انتخاب شده است.

۵ می‌باشد [۱۴]. بنابراین ما مدل را ۵ بار متمایز بر روی ۸۰٪ داده‌ها آموزش دادیم و بر روی باقیمانده ۲۰٪ داده‌ها آزمایش نمودیم. در این پروژه در ابتدا مدل پیشنهادی را برای هر فرد بصورت جداگانه آموزش داده‌ایم و عملکرد مدل را بر روی داده‌های آزمایشی همان فرد ارزیابی نمودیم. سپس در مرحله بعدی، مدل پیشنهادی را بر روی داده‌های آموزشی تمامی افراد بصورت همزمان آموزش دادیم و عملکرد آن را در دسته بندی داده‌های آزمایشی هر فرد و همچنین تمام مجموعه داده ارزیابی نمودیم.

۵- مدل‌های پیشنهادی

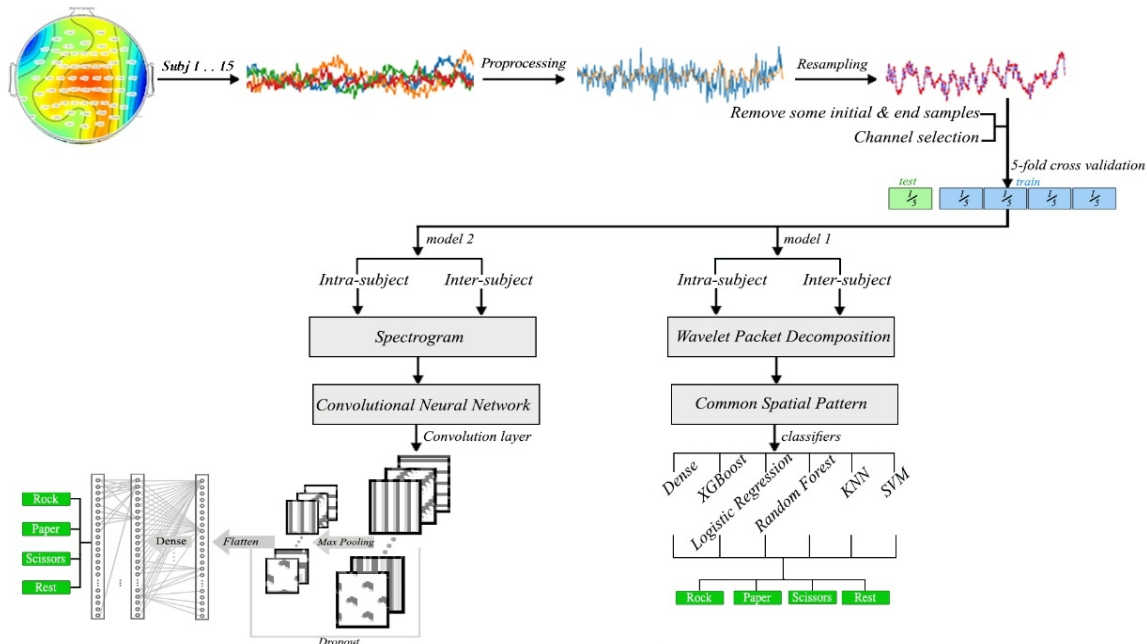
ما در این بخش مدل‌های مورد استفاده در این پژوهش را توضیح خواهیم داد. مراحل پردازش و نحوه عملکرد مدل‌های اولیه و توسعه یافته در شکل ۶ نشان داده شده است.

۱-۵- مدل اول

با استفاده از این مدل ما توانستیم مقام دوم در مسابقه سومین مسابقه ملی واسط مغز و رایانه را کسب نماییم. این مدل از استخراج و انتخاب ویژگی‌ها با استفاده از تجزیه بسته‌های موجک (WPD) و الگوی مکانی مشترک (CSP) تشکیل شده است. سپس عملکرد مدل توسط دسته‌بندی‌های مختلف مورد ارزیابی قرار گرفت.

۱-۱-۵- تجزیه بسته‌های موجک (WPD)

تجزیه بسته‌های موجک، حالت پیشرفته‌تر تبدیل موجک گسسته (DWT) است به نحوی که سیگنال از فیلترهای بیشتری عبور داده می‌شود. در تبدیل موجک گسسته در هر مرحله، ضریب تقریب از فیلترهای پایین‌گذر و بالاگذر زمان گسسته عبور داده می‌شود، ولی در



شکل ۶: مراحل پردازش و نحوه عملکرد مدل‌های پیشنهادی

۳-۱-۵- دسته‌بندها

ما دسته‌بندهای ماشین بردار پشتیبان (SVM)، k -نزدیک‌ترین همسایه (KNN)، جنگل تصادفی (RF)، رگرسیون لجستیک (LR)، XGBoost و مدل یادگیری عمیق Dense را بر روی ویژگی‌های بدست آمده پیاده‌سازی نمودیم. با استفاده از الگوریتم Grid Search پارامترهای بهینه برای هر یک از دسته‌بندها مورد جستجو و ارزیابی قرار گرفت. عمق جنگل در RF برابر با ۳ و تعداد درختان برابر با ۶۵ در نظر گرفته شد. تعداد همسایه در KNN برابر با ۱۶ انتخاب گردید. برای SVM کرنل‌های خطی و RBF بطور مشابه بهترین عملکرد را داشتند. در LR از میان solverهای مختلف، بهترین جواب برای SAG بوده است. در XGBoost بهترین عملکرد برای نرخ یادگیری ۰/۰۵، تعداد تخمین‌زندگان ۵۰۰ با عمق ۸ بوده است. برای مدل Dense از دو لایه Dense با ۱۲۴ نورون در هر لایه و تابع فعالساز Relu و Dropout برابر با ۰/۵ استفاده شده است. همچنین لایه Dense آخر با ۴ نورون با تابع فعالساز Softmax، تابع بهینه‌ساز Adam و تابع هزینه Categorical_crossentropy برای دسته بندی نهایی بکار گرفته شده است.

در مدل Dense از توابع Callback در مرحله آموزش برای کنترل نحوه روزرسانی وضعیت‌ها در حین آموزش استفاده نمودیم. از EarlyStopping برای توقف آموزش در صورتی که دقت مدل در ۵۰ دوره (epoch) متوالی بهبود نیافت استفاده نمودیم. همچنین از ReduceLROnPlateau برای کاهش نرخ یادگیری با ضریب ۰/۱ در صورت توقف پیشرفت مدل در ۱۵ دوره متوالی استفاده نمودیم. با استفاده از ModelCheckpoint، وزن‌های اصلاح‌شده مدل در صورتی که دقت مدل نسبت به دوره قبل افزایش یافته باشد ذخیره گردیدند.

عملکرد مدل در هر دو حالت آموزش/آزمایش یک فرد، و آموزش/آزمایش همزمان تمامی افراد مورد بررسی قرار گرفت. نتایج حاصل در جدول شماره ۱ نشان داده شده است. نتایج حاصل، از میانگین دقت های بدست آمده از ۵ بار اعتبارسنجی متقابل به دست آمده است. همانگونه که در نتایج مشاهده می شود دسته‌بند XGBoost بهترین دقت میانگین را برای آموزش جداگانه افراد داشته است در حالی که دسته‌بند Dense بهترین دقت میانگین را برای آموزش همزمان تمامی افراد داشته است. مقایسه عملکرد دسته‌بندهای مختلف به تفکیک هر فرد در دو حالت آموزش جداگانه برای هر فرد و آموزش همزمان برای تمامی افراد در شکل ۷ نشان داده شده است. همچنین میانگین دقت دسته‌بندهای مختلف در شکل ۸ با یکدیگر مقایسه شده‌اند.

۲-۵- مدل توسعه داده شده

در این مدل که به منظور رسیدن به عملکرد بهتری در مقایسه با مدل اول توسعه داده شد از روش اسپکتروگرام در کنار شبکه عصبی کانولوشنی استفاده نمودیم.

۱-۲-۵- اسپکتروگرام

اسپکتروگرام نمایشی از طیف سیگنال‌هایی است که نسبت به زمان تغییرات زیادی دارند. اسپکتروگرام معمولاً با یکی از دو روش تقریب‌زدن به صورت یک بانک فیلتری (مجموعه‌ای از فیلترهای میان‌گذر) و یا استفاده از تبدیل فوریه در حوزه زمان ساخته می‌شود. روش اول معمولاً در پردازش‌های آنالوگ مورد استفاده قرار می‌گیرد و سیگنال ورودی را به باندهای فرکانسی مختلف تبدیل می‌کند. روش دوم در پردازش‌های دیجیتال مورد بهره‌برداری قرار می‌گیرد و از تبدیل فوریه در حوزه زمان بهره می‌برد. یکی از معایب تبدیل فوریه کلاسیک عدم انعکاس رفتارهای محلی سیگنال می‌باشد. یک راه ساده برای غلبه بر این مشکل، مقایسه سیگنال با توابع پایه‌ای است که در حوزه‌های زمان و فرکانس توانمند محلی شده باشند. تبدیل فوریه کوتاه‌مدت یا به اختصار STFT یکی از روش‌های اولیه مقایسه سیگنال با این توابع مقدماتی بوده و به صورت زیر تعریف می‌شود:

$$STFT(t, \omega) = \int_{-\infty}^{+\infty} s(t) \gamma_{t, \omega}^*(t) dt \quad (1)$$

در رابطه بالا، $\gamma(t)$ یک تابع زمانی دارای طول محدود به نام تابع پنجره می‌باشد. تبدیل فوریه کوتاه‌مدت به عبارتی تبدیل فوریه پنجره‌شده سیگنال در فواصل زمانی مختلف می‌باشد. مجذور دامنه یا اندازه STFT را اسپکتروگرام گویند.

$$Spectrogram: SP(t, \omega) = |STFT(t, \omega)|^2 \quad (2)$$

در اولین مرحله، سیگنال پیش‌پردازش شده به اسپکتروگرام با طول محدود تبدیل می‌شود. این تبدیل با استفاده از الگوریتم استاندارد تبدیل سریع فوریه توسط پنجره‌های همپینگ با همپوشانی صورت گرفته و در کنار یکدیگر قرار داده می‌شود. در پایان این مرحله، سیگنال یک‌بعدی به تصویری دوبعدی تبدیل می‌شود.

۲-۲-۵- شبکه عصبی کانولوشنی

شبکه عصبی کانولوشنی (CNN) یک تصویر را به عنوان ورودی دریافت نموده و ویژگی‌های آن را با استفاده از لایه‌های کانولوشن و Pooling استخراج می‌نماید. سپس ویژگی‌های استخراج‌شده به لایه‌های Flattening و Dense به منظور دسته‌بندی وارد می‌شود. لایه‌های پشت سرهم این شبکه در ادامه توضیح داده شده است [۱۶].

الف) لایه کانولوشن: یک تصویر را به عنوان ورودی دریافت می‌نماید و ویژگی‌های موردنظر را با استفاده از روش‌های پردازش تصویر نظیر هموارسازی، حذف نویز، و تشخیص لبه استخراج می‌نماید.

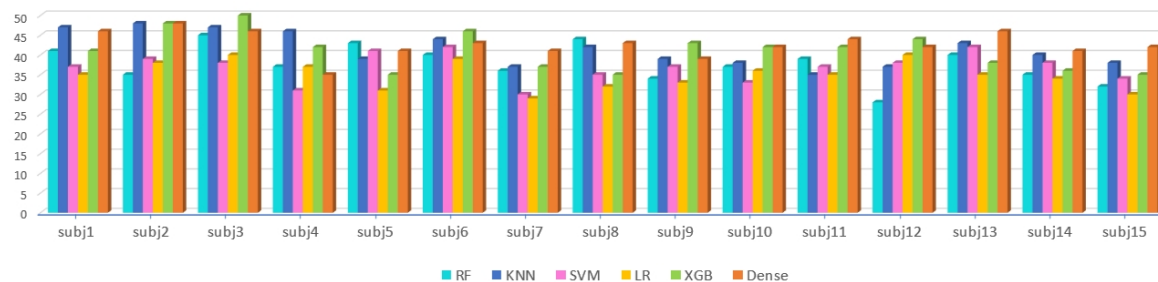
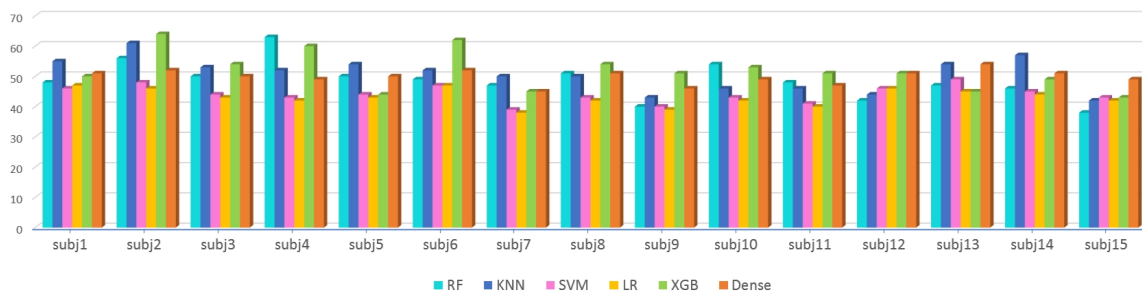
ب) لایه Pooling: ابعاد ماتریس ویژگی‌ها را با حفظ ویژگی‌های مهم کاهش می‌دهد.

ج) لایه Flattening: ماتریس ویژگی‌ها را به عنوان ورودی دریافت نموده و به یک بردار تک‌بعدی تبدیل می‌نماید.

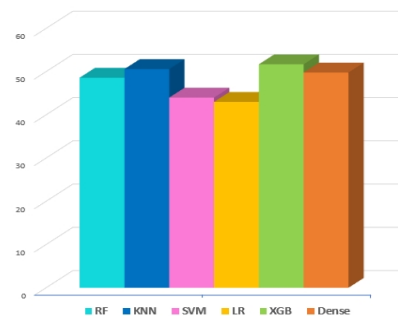
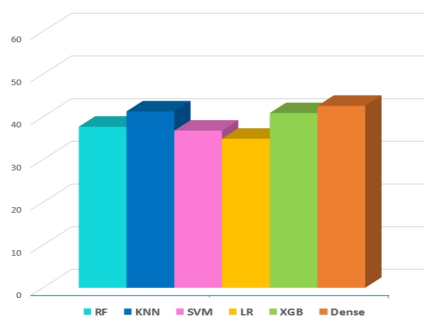
د) لایه Dense: بردار تک‌بعدی حاصل را به یک یا چند لایه عصبی به منظور انجام دسته‌بندی می‌دهد.

جدول ۱: عملکرد دسته بندهای مختلف به تفکیک هر فرد در دو حالت آموزش جداگانه و همزمان

آموزش	subj1	subj2	subj3	subj4	subj5	subj6	subj7	subj8	subj9	subj10	subj11	subj12	subj13	subj14	subj15	میانگین		
RF	جداگانه	48%	56%	50%	63%	50%	49%	47%	51%	40%	54%	48%	42%	47%	46%	38%	48.6%	
	همزمان	41%	35%	45%	37%	43%	40%	36%	44%	34%	37%	39%	28%	40%	35%	32%	37.7%	
KNN	جداگانه	55%	61%	53%	52%	54%	52%	50%	43%	46%	46%	44%	54%	57%	42%	50.6%		
	همزمان	47%	48%	47%	46%	39%	44%	37%	42%	39%	38%	35%	37%	43%	40%	38%	41.3%	
SVM	جداگانه	46%	48%	44%	43%	44%	47%	39%	43%	40%	43%	41%	46%	49%	45%	43%	44.0%	
	همزمان	37%	39%	38%	31%	41%	42%	30%	35%	37%	33%	37%	38%	42%	38%	34%	36.8%	
LR	جداگانه	47%	46%	43%	42%	43%	47%	38%	42%	39%	42%	40%	46%	45%	44%	42%	43.0%	
	همزمان	35%	38%	40%	37%	38%	29%	32%	33%	36%	40%	35%	40%	35%	34%	30%	34.9%	
XGB	جداگانه	50%	64%	54%	60%	44%	62%	45%	54%	51%	53%	51%	51%	45%	49%	43%	51.7%	
	همزمان	41%	48%	50%	42%	35%	46%	37%	35%	43%	42%	42%	44%	38%	36%	35%	40.9%	
Dense	جداگانه	51%	52%	50%	49%	50%	52%	45%	51%	46%	49%	47%	51%	54%	51%	49%	49.8%	
	همزمان	46%	48%	46%	35%	41%	43%	41%	43%	39%	42%	44%	42%	46%	41%	42%	42.6%	



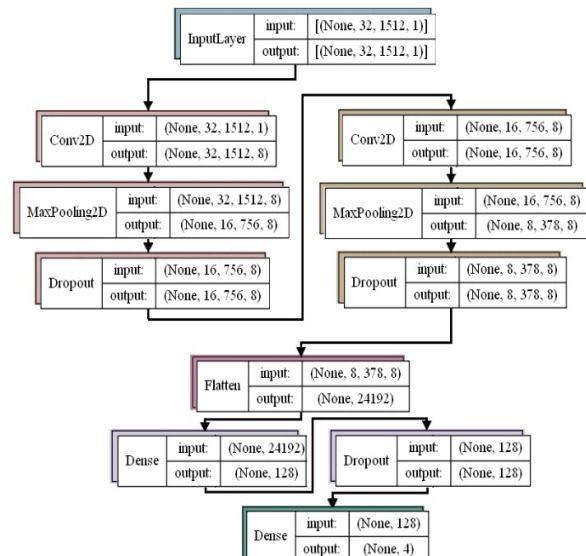
شکل ۷: مقایسه عملکرد دسته بندهای مختلف به تفکیک هر فرد در مدل پیشنهادی اول



شکل ۸: مقایسه میانگین دقت دسته بندهای مختلف در مدل پیشنهادی اول

۳-۲-۵- معماری مدل

ما یک مدل مبتنی بر CNN دو بعدی شامل ۱۰ لایه ایجاد نمودیم. این مدل از ۲ لایه کانولوشنی، ۲ لایه Pooling، ۱ لایه Flattening، ۲ لایه Dense و ۳ لایه Dropout تشکیل شده است. Dropout روشی برای جلوگیری از بیش‌برازش (overfitting) در هنگام آموزش شبکه می‌باشد. در این روش بصورت تصادفی مقادیر برخی از نورون‌های لایه‌های مخفی شبکه صفر می‌شود. بنابراین لایه‌های مخفی قادر به بخاطر سپردن تمامی جزئیات نمونه‌های آموزشی و در نتیجه بیش‌برازش شبکه نخواهند بود. معماری CNN مورد استفاده در شکل ۹ نشان داده شده است.



شکل ۹: معماری شبکه عصبی کانولوشنی مورد استفاده

۴-۲-۵- تنظیمات ابرپارامترها

به دلیل وجود چهار دسته، دسته‌بند مورد استفاده بصورت چنددسته‌ای بوده و از Softmax به عنوان تابع فعال‌سازی لایه آخر استفاده نمودیم. توابع فعال‌سازی لایه‌های پیشین Relu بوده و از Sparse_categorical_crossentropy به عنوان تابع هزینه استفاده نمودیم. Dropout برابر با ۰/۲ تنظیم شده است. ما همچنین بهترین بهینه‌ساز را برای مدل از میان بهینه‌سازهای SGD، Adam و RMSprop و بهترین نرخ یادگیری را از میان نرخ‌های ۰/۱، ۰/۰۱، ۰/۰۰۱ و ۰/۰۰۰۱ با استفاده از الگوریتم Grid Search جستجو نمودیم. بهترین نتیجه برای مدل با استفاده از بهینه‌ساز Adam و نرخ یادگیری ۰/۰۰۰۱ به دست آمد.

ما همچنین از توابع Callback در مرحله آموزش برای کنترل نحوه روزرسانی وضعیت‌ها درحین آموزش استفاده نمودیم. از EarlyStopping برای توقف آموزش در صورتی که دقت مدل در ۵۰ دوره (epoch) متوالی بهبود نیافت استفاده نمودیم. همچنین از ReduceLROnPlateau برای کاهش نرخ یادگیری با ضریب ۰/۱ در صورت توقف پیشرفت مدل در ۱۵ دوره متوالی استفاده نمودیم. با

استفاده از ModelCheckpoint، وزن‌های اصلاح‌شده مدل در صورتی که دقت مدل نسبت به دوره قبل افزایش یافته باشد ذخیره گردیدند.

۵-۲-۵- نتایج عملکرد مدل

عملکرد مدل توسعه یافته بر روی داده‌های آزمایشی در دو حالت آموزش جداگانه و آموزش همزمان به تفکیک هر فرد در جدول ۲ نشان داده شده است. دقت‌های درج شده از میانگین دقت‌های ۵ بار اعتبارسنجی متقابل به دست آمده است. همچنین ماتریس درهم‌ریختگی نتایج دسته‌بندی دسته‌ها به تفکیک هر فرد برای آموزش جداگانه در شکل ۱۰ نشان داده شده است. ماتریس درهم‌ریختگی عملکرد کلی مدل در دسته‌بندی دسته‌ها برای هر دو حالت جداگانه و همزمان در شکل ۱۱ نشان داده شده است.

۶- بحث و نتیجه گیری

استخراج و رمزگشایی صحبت‌های ذهنی انسان علاوه بر توانمندسازی افراد دچار مشکلات تکلم، دریچه جدیدی بر روی ارتباطات و تعاملات انسان‌ها می‌گشاید. مجموعه داده این پژوهش توسط مرکز ملی نقشه‌برداری مغز ایران برای سومین مسابقه ملی واسط مغز و رایانه منتشر شده است و شامل سیگنال‌های EEG از ۱۵ شرکت‌کننده می‌باشد که کلمات سنگ، کاغذ و قیچی را در ذهن خویش بیان می‌نمایند. همچنین سیگنال‌های زمان استراحت نیز در مجموعه داده موجود می‌باشد. ما در ابتدا داده‌ها را به منظور حذف نویز، آرتیفکت و فرکانس‌های نامربوط پیش پردازش نمودیم. سپس با کاهش نرخ نمونه برداری، امکان حذف افزونگی داده‌ها و کاهش زمان پردازش و هزینه محاسباتی را ممکن ساختیم. روش اعتبارسنجی متقابل ۵ بخشی را برای تفکیک داده‌ها به منظور ارزیابی دقیق‌تر عملکرد مدل استفاده نمودیم.

در ابتدا مدل پیشنهادی را برای هر فرد بصورت جداگانه آموزش دادیم و عملکرد مدل را بر روی داده‌های آزمایشی همان فرد ارزیابی نمودیم. سپس در مرحله بعدی، مدل پیشنهادی را بر روی داده‌های آموزشی تمامی افراد بصورت همزمان آموزش دادیم و عملکرد آن را در دسته بندی داده‌های آزمایشی هر فرد و تمام مجموعه داده‌ها ارزیابی نمودیم. مدل مورد استفاده ما در مسابقه از ترکیب دو روش تجزیه بسته‌های موجک و الگوی مکانی مشترک برای استخراج و انتخاب ویژگی‌های مطلوب بهره برده است و سپس با استفاده از دسته‌بندی ماشین بردار پشتیبان، k_ نزدیک‌ترین همسایه، جنگل تصادفی، رگرسیون لجستیک، XGBoost و یادگیری عمیق Dense سعی نمودیم بهترین دقت مدل را در دسته‌بندی چهار دسته سنگ، کاغذ، قیچی و استراحت بدست آوریم. دسته بند XGBoost بهترین دقت میانگین را برای آموزش جداگانه افراد داشته است (۵۱/۷٪) در حالی که دسته‌بند Dense بهترین دقت میانگین را برای آموزش همزمان تمامی افراد داشته است (۴۲/۶٪). بر اساس دقت های بدست آمده، آموزش دسته‌بند بطور جداگانه توسط داده‌های آموزشی هر فرد، افزایش دقت قابل توجهی نسبت به آموزش دسته‌بند بطور همزمان توسط داده‌های آموزشی تمام افراد دارد.

جدول ۲: دقت مدل توسعه یافته در دو حالت آموزش جداگانه و همزمان به تفکیک هر فرد

میانگین	subj15	subj14	subj13	subj12	subj11	subj10	subj9	subj8	subj7	subj6	subj5	subj4	subj3	subj2	subj1	اعتبارسنجی متقابل ۵ بخشی	آموزش	
87%	80%	85%	85%	85%	100%	95%	70%	95%	90%	70%	100%	85%	100%	90%	75%	بهترین	جداگانه	CNN
65.6%	65%	65%	50%	65%	80%	80%	50%	75%	55%	50%	85%	70%	70%	70%	55%	بدترین		
76.5%	73%	74%	66%	79%	89%	84%	59%	87%	72%	61%	93%	77%	86%	83%	65%	میانگین		
74.3%	90%	80%	85%	75%	85%	95%	70%	75%	65%	60%	70%	70%	60%	70%	65%	بهترین	همزمان	
56.6%	70%	60%	50%	65%	60%	70%	65%	60%	55%	50%	40%	65%	35%	60%	45%	بدترین		
66%	81%	72%	67%	71%	76%	79%	69%	69%	59%	54%	59%	66%	48%	65%	56%	میانگین		

Actual	Rest	1 4%	0	0	24 96%
	Scissors	2 8%	10 40%	10 40%	3 12%
	Paper	0	18 72%	1 4%	6 24%
	Rock	13 52%	4 16%	3 12%	5 20%
RockPaperScissors Rest					
Predicted					
Subject 1					
Actual	Rest	0	0	0	25 100%
	Scissors	1 4%	2 8%	22 88%	0
	Paper	1 4%	19 76%	4 16%	1 4%
	Rock	17 68%	2 8%	4 16%	2 8%
RockPaperScissors Rest					
Predicted					
Subject 2					
Actual	Rest	1 4%	0	0	24 96%
	Scissors	0	0	24 96%	1 4%
	Paper	2 8%	22 88%	0	1 4%
	Rock	16 64%	4 16%	2 8%	3 12%
RockPaperScissors Rest					
Predicted					
Subject 3					
Actual	Rest	3 12%	0	0	22 88%
	Scissors	5 20%	4 16%	15 60%	1 4%
	Paper	5 20%	18 72%	2 8%	0
	Rock	22 88%	2 8%	1 4%	0
RockPaperScissors Rest					
Predicted					
Subject 4					
Actual	Rest	0	0	0	25 100%
	Scissors	0	3 12%	20 80%	2 8%
	Paper	0	18 72%	3 12%	4 16%
	Rock	10 40%	4 16%	7 28%	4 16%
RockPaperScissors Rest					
Predicted					
Subject 5					
Actual	Rest	0	0	0	25 100%
	Scissors	0	2 8%	22 88%	1 4%
	Paper	0	24 96%	1 4%	0
	Rock	23 92%	1 4%	0	1 4%
RockPaperScissors Rest					
Predicted					
Subject 6					
Actual	Rest	0	1 4%	0	24 96%
	Scissors	1 4%	3 12%	20 80%	2 8%
	Paper	0	18 72%	3 12%	4 16%
	Rock	10 40%	4 16%	7 28%	4 16%
RockPaperScissors Rest					
Predicted					
Subject 7					
Actual	Rest	0	0	0	25 100%
	Scissors	0	3 12%	15 60%	7 28%
	Paper	0	13 52%	0	12 48%
	Rock	8 32%	1 4%	3 12%	13 52%
RockPaperScissors Rest					
Predicted					
Subject 8					
Actual	Rest	0	0	0	25 100%
	Scissors	2 8%	0	22 88%	1 4%
	Paper	1 4%	18 72%	6 24%	0
	Rock	22 88%	1 4%	2 8%	0
RockPaperScissors Rest					
Predicted					
Subject 9					
Actual	Rest	0	0	0	25 100%
	Scissors	3 12%	1 4%	9 36%	12 48%
	Paper	3 12%	13 52%	1 4%	8 32%
	Rock	12 48%	0	4 16%	9 36%
RockPaperScissors Rest					
Predicted					
Subject 10					
Actual	Rest	1 4%	0	0	24 96%
	Scissors	1 4%	3 12%	20 80%	1 4%
	Paper	0	24 96%	1 4%	0
	Rock	21 84%	3 12%	1 4%	0
RockPaperScissors Rest					
Predicted					
Subject 11					
Actual	Rest	0	0	1 4%	24 96%
	Scissors	1 4%	1 4%	21 84%	2 8%
	Paper	3 12%	17 68%	0	5 20%
	Rock	17 68%	3 12%	1 4%	4 16%
RockPaperScissors Rest					
Predicted					
Subject 12					
Actual	Rest	1 4%	0	0	24 96%
	Scissors	2 8%	1 4%	17 68%	5 20%
	Paper	2 8%	15 60%	0	8 32%
	Rock	17 68%	1 4%	1 4%	6 24%
RockPaperScissors Rest					
Predicted					
Subject 13					
Actual	Rest	0	0	0	25 100%
	Scissors	4 16%	2 8%	12 48%	7 28%
	Paper	1 4%	22 88%	0	2 8%
	Rock	15 60%	8 32%	0	2 8%
RockPaperScissors Rest					
Predicted					
Subject 14					
Actual	Rest	1 4%	0	0	24 96%
	Scissors	2 8%	1 4%	17 68%	5 20%
	Paper	2 8%	15 60%	0	8 32%
	Rock	17 68%	1 4%	1 4%	6 24%
RockPaperScissors Rest					
Predicted					
Subject 15					

شکل ۱۰: ماتریس درهم ریختگی چگونگی دسته بندی دسته ها به تفکیک هر فرد برای حالت آموزش جداگانه

Actual	Rest	8 2.1%	2 0.6%	4 1%	361 96.2%
	Scissors	25 6.6%	38 10.1%	260 69.3%	52 13.9%
	Paper	23 6.13%	273 72.8%	24 6.4%	55 14.7%
	Rock	254 67.7%	35 9.4%	33 8.8%	53 14.1%
		Rock	Paper	Scissors	Rest

Predicted

(الف) آموزش جداگانه

Actual	Rest	16 4.2%	2 0.6%	6 1.6%	351 93.6%
	Scissors	38 10.1%	77 20.6%	195 52%	65 17.3%
	Paper	30 8%	249 66.4%	27 7.2%	69 18.4%
	Rock	196 52.2%	71 18.9%	31 8.3%	77 20.5%
		Rock	Paper	Scissors	Rest

Predicted

(ب) آموزش همزمان

شکل ۱۱: ماتریس درهم ریختگی عملکرد کلی مدل در نحوه دسته بندی دسته ها

از طرف دیگر مشاهده می شود کمترین خطای دسته بندی نیز دسته بندی اشتباه حالت استراحت بجای کاغذ با درصد خطای ۰/۶٪ می باشد. همچنین دسته بندی اشتباه حالت استراحت به عنوان سنگ و قیچی به ترتیب برابر با ۲/۱٪ و ۱٪ است که نشان می دهد سیگنال استراحت تفاوت معناداری با سیگنال های دسته های دیگر دارد. لذا دسته بندی اشتباه دسته های دیگر به عنوان حالت استراحت، نتیجه عدم وجود داده های آموزشی کافی برای هر کدام از این دسته ها می باشد. به دلیل تفاوت سیگنال های مغزی در افراد مختلف و حتی در یک فرد در ثبت های مختلف، وجود داده های آموزشی کافی برای هر فرد ضروری می باشد. با توجه به وجود تنها ۱۰۰ نمونه ثبت شده از هر فرد که ۸۰ نمونه آن صرف آموزش مدل می شود تفاوت سیگنال های ثبت شده متعلق به یک دسته می تواند سبب شود که سیگنال های آزمایشی یک دسته متفاوت از سیگنال های آموزشی آن دسته باشد. به دلیل این که سیگنال های حالت استراحت به دلیل وجود نوسان کم بسیار شبیه به یکدیگر و دارای فراوانی و تکرارپذیری بالا هستند سیگنال های ناشناخته دسته های دیگر به اشتباه به حالت استراحت که تکرارپذیرتر می باشد دسته بندی می شوند.

همچنین مشاهده می شود که بیشترین خطای دسته بندی متقابل بین سه دسته سنگ، کاغذ و قیچی، به تشخیص اشتباه قیچی به عنوان کاغذ با ۱۰/۱٪ خطا اختصاص دارد. تشخیص اشتباه سنگ به عنوان کاغذ ۹/۴٪، سنگ به عنوان قیچی ۸/۸٪، قیچی به عنوان سنگ ۶/۶٪، کاغذ به عنوان قیچی ۶/۴٪ و کاغذ به عنوان سنگ ۶/۱۳٪ خطا را شامل می شوند. مشاهده می شود که درصد خطاهای دسته بندی این سه دسته نزدیک به یکدیگر بوده و هیچ یک از دو دسته دارای رابطه متقابل معناداری با یکدیگر نمی باشند. در جدول ۳ دقت مدل توسعه یافته با تعدادی از پژوهش های جدید در این زمینه مقایسه شده است.

همانگونه که در جدول ۳ مشاهده می شود مدل پیشنهادی ما در مقایسه با پژوهش های پیشین عملکردی بهتر داشته است. همچنین در تنها پژوهش [۱۱] منتشر شده بر روی این مجموعه داده، که مقام اول مسابقات را نیز کسب نموده است دقت مدل ۵۱/۹٪ گزارش شده است. بهترین دقت مدل اولیه ما که در مسابقه ارزیابی شده بود با استفاده از روش های تجزیه بسته های موجک، الگوی مکانی مشترک و دسته بندی

ما در ادامه این مدل را توسعه دادیم و توانستیم به دقتی بسیار بهتر از بالاترین دقتی که تاکنون بر روی این مجموعه داده گزارش شده است دست یابیم. مدل توسعه داده شده از اسپکتروگرام برای تبدیل داده ها بهره می برد و سپس با استفاده از یادگیری عمیق و شبکه عصبی کانولوشنی به استخراج ویژگی های مطلوب و دسته بندی می پردازد. میانگین دقت بدست آمده بر روی مجموعه داده های آموزشی در این مدل برای آموزش جداگانه برابر با ۷۶/۵٪ و برای آموزش همزمان برابر با ۶۶٪ بدست آمده است. هر دو دقت بدست آمده در مقایسه با مدل اول ارتقای مطلوبی داشته اند. همچنین از بهترین دقت بدست آمده در مسابقه توسط نفر اول که برابر با ۵۱/۹٪ بوده است نتیجه بسیار بهتری داشته است. در این مدل نیز مشاهده می شود که آموزش جداگانه بر روی داده های هر فرد، عملکرد بهتری نسبت به آموزش همزمان بر روی داده های تمامی افراد داشته است. این به دلیل تفاوت ویژگی سیگنال های مغزی افراد با یکدیگر و همچنین تفاوت در تصور بیان کلمه توسط افراد مختلف است.

همانگونه که از ماتریس های درهم ریختگی میانگین مشاهده می شود بهترین دسته بندی صحیح برای دسته استراحت با ۹۶/۲٪ صحت دسته بندی می باشد. تصور بیان کلمات کاغذ با ۷۲/۸٪، قیچی با ۶۹/۳٪، و سنگ با ۶۷/۷٪ رتبه های بعدی دسته بندی صحیح را تشکیل داده اند. به دلیل این که سیگنال های مغزی در حالت استراحت دارای نوسان کم هستند تشخیص آن ها از سیگنال های در حالت تمرکز و فعالیت راحت تر می باشد. دسته بندی اشتباه کاغذ به عنوان استراحت با ۱۴/۷٪ بیشترین خطای دسته بندی را شامل می شود. از ماتریس درهم ریختگی مشاهده می شود که بیشترین خطاها، دسته بندی اشتباه سنگ، کاغذ و قیچی به عنوان حالت استراحت به ترتیب برابر با ۱۴/۷٪، ۱۴/۹٪ و ۱۳/۹٪ خطا می باشد. این می تواند به دلیل وجود حالت استراحت در سیگنال های سنگ، کاغذ و قیچی پیش یا پس از اتمام تصور بیان باشد. همچنین در صورتی که نویزهای ناشی از عملکرد نامناسب الکترودها، نویزهای محیطی، و آرتیفیکت های ماهیچه ای اثر یکسان بر روی تمامی سیگنال ها در حالت تصور بیان و در حالت استراحت داشته باشند سیگنال های دسته های مختلف دارای شباهت نزدیک می گردند. اما دقت بالای مدل در دسته بندی صحیح هر دسته، نشان از کیفیت خوب سیگنال ها دارد.

جدول ۳: مقایسه دقت مدل پیشنهادی با تعدادی از پژوهش های جدید در این زمینه

پژوهش	سال	مجموعه داده	روش استخراج ویژگی ها	دسته بند	بهترین دقت میانگین (%)
این پژوهش	۲۰۲۱	کلمات سنگ، کاغذ، قیچی با ۱۵ شرکت کننده	تجزیه بسته های موجک الگوی مکانی مشترک اسپکتروگرام	SVM KNN Random Forest Logistic Regression XGBoost Dense CNN	۷۶/۵
پژوهش [۸]	۲۰۲۱	چهار کلمه اسپانیایی با ۲۰ شرکت کننده	تجزیه موجک تئوری آشوب بعد فرکتال انرژی فرکانس تجزیه مولفه های اصلی	SVM KNN Random Forest Logistic Regression	۶۵
پژوهش [۹]	۲۰۲۰	شش کلمه اسپانیایی با ۱۵ شرکت کننده	یادگیری عمیق	Shallow CNN CNN EEGNet	۲۴/۹۰
پژوهش [۱۰]	۲۰۲۱	۵ کلمه اسپانیایی با ۱۵ شرکت کننده	ترکیب یادگیری عمیق تطبیق دامنه بدون ناظر	Standardization-refinement domain adaptation	۶۲/۹۹
پژوهش [۱۱]	۲۰۲۰	همانند این مقاله	تجزیه بسته های موجک الگوی مکانی مشترک	XGBoost	۵۱/۷

- [3] Y. Wang, M. Zhang, R. Wu, H. Gao, M. Yang, Z. Luo, G. Li, Silent speech decoding using spectrogram features based on neuromuscular activities, Brain Sciences 10 (2020) 442.
- [4] A.R. Sereshkeh, R. Yousefi, A.T. Wong, T. Chau, Online classification of imagined speech using functional near-infrared spectroscopy signals, Journal of neural engineering 16 (2018) 016005.
- [5] S. Martin, I. Iturrate, P. Brunner, J.d.R. Millán, G. Schalk, R.T. Knight, B.N. Pasley, Individual Word Classification During Imagined Speech Using Intracranial Recordings, Brain-Computer Interface Research, Springer 2019, pp. 83-91.
- [6] A. Aleman, E. Formisano, H. Koppenhagen, P. Hagoort, E.H. De Haan, R.S. Kahn, The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words, Cerebral Cortex 15 (2005) 221-228.
- [7] L. Wang, X. Zhang, X. Zhong, Y. Zhang, Analysis and classification of speech imagery EEG for BCI, Biomedical signal processing and control 8 (2013) 901-908.
- [8] T. Hernández-Del-Toro, C.A. Reyes-García, L. Villaseñor-Pineda, Toward asynchronous EEG-based BCI: Detecting imagined words segments in continuous EEG signals, Biomedical Signal Processing and Control 65 (2021) 102351.
- [9] C. Cooney, A. Korik, R. Folli, D. Coyle, Evaluation of Hyperparameter Optimization in Machine and Deep Learning Methods for Decoding Imagined Speech EEG, Sensors 20 (2020) 4629.
- [10] M. Jiménez-Guarneros, P. Gómez-Gil, Standardization-refinement domain adaptation method for cross-subject EEG-based classification in imagined speech recognition, Pattern Recognition Letters 141 (2021).
- [11] N. Hamed, S. Samiei, M. Delrobaei, A. Khadem, Imagined Speech Decoding From EEG: The Winner of 3rd Iranian BCI Competition (iBCIC2020), 2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME), IEEE, 2020.
- [12] <https://nbml.ir/FA/scientific-tournament/102640/> (accessed 13 October 2020).
- [13] Makoto's preprocessing pipeline, https://scn.ucsd.edu/wiki/Makoto's_preprocessing_pipeline (accessed 25 September 2019).
- [14] T. Fushiki, Estimation of prediction error by using K-fold cross-validation, Statistics and Computing 21 (2011) 137-146.
- [15] H. Ramoser, J. Muller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, IEEE transactions on rehabilitation engineering 8 (2000) 441-446.
- [16] M. Moghaddari, M.Z. Lighvan, S. Danishvar, Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG, Computer Methods and Programs in Biomedicine 197 (2020).

XGBoost برابر با ۵۱/۷٪ به دست آمده بود که اختلاف ناچیزی با دقت مقام اول مسابقه داشته است. مدل توسعه داده شده ما با میانگین دقت ۷۶/۵٪ توانست میانگین دقت پیشین مدل را ۲۴/۸٪ ارتقا دهد.

با توجه به این که تشخیص تصویر گفتار برای کلمات فارسی تاکنون بندرت مورد پژوهش قرار گرفته است در این پژوهش سعی شده است برای اولین بار روش ها و مدل های مختلفی در تحلیل دادگان تصور گفتار واژگان فارسی بررسی شود و مقایسه دقیقی بین عملکرد مدل های مختلف انجام پذیرد. نتایج حاصل نشان داده است که امکان استفاده از کلمات فارسی برای تحلیل تصور گفتار وجود دارد و رمزگشایی از تصور گفتار محدود به زبان های بین المللی پیشتر تحقیق شده نیست. همچنین مدلی که بهترین عملکرد را بر روی دادگان موجود داشته است معرفی گردید. با توجه به دسته بندها و مدل های متعدد بررسی شده، این پژوهش همچنین می تواند به عنوان یک راهنما برای پژوهش های آینده مورد استفاده قرار گیرد.

با توجه به جذابیت و جدید بودن موضوع رمزگشایی از تصور گفتار بخصوص برای واژگان فارسی، و همچنین به منظور کاربردی نمودن ابزارهای تصور گفتار نیاز است مدل پیشنهادی برای دستیابی به نتایج بهتر و کاربردی تر، توسعه داده شود. همچنین پیشنهاد می شود مدل توسعه داده شده بر روی مجموعه داده های بزرگتری با تعداد شرکت کنندگان بیشتر و همچنین مجموعه کلمات بیشتر ارزیابی شود. پس از تشخیص تصور بیان کلمات، تشخیص تصور بیان جملات می تواند گامی موثر در عملیاتی نمودن سیستم های ارتباطات تنها با صحبت ذهنی بین افراد باشد.

مراجع

- [1] P.R. Roelfsema, D. Denys, P.C. Klink, Mind reading and writing: The future of neurotechnology, Trends in cognitive sciences 22 (2018)
- [2] D. Dash, P. Ferrari, J. Wang, Decoding imagined and spoken phrases from non-invasive neural (MEG) signals, Frontiers in neuroscience 14 (2020).