دانشگاه شهید بهشتی
دانشکده مهندسی و علوم کامپیوتر

بررسی روش‌های مدل‌سازی مبتنی بر شبکه عصبی برای دسته‌بندی تصورات کلمات و واج‌ها از روی سیگنال مغزی EEG

پروژه کارشناسی مهندسی کامپیوتر

دانشجو:
عرفان قبادیان

استاد راهنما:
دکتر یاسر شکفته

زمستان ۱۴۰۱

چکیده

صحبت کردن مکانیزمی پیچیده است که بخش‌های مختلف مغز را در روند تولید، برنامه‌ریزی و کنترل دقیق تعداد زیادی از عضلات مربوط به واجگاه‌ها، برای ادا کردن کلمات و واج‌ها و در نهایت ساخت جملات درگیر می‌کند. در عین حال صحبت کردن یکی از مهم‌ترین راه‌های ارتباط انسان‌ها با یکدیگر است. برخی افراد به دلیل بیماری و اختلالات مختلف قادر به صحبت کردن نیستند. برای ساده کردن راه‌های ارتباط این افراد، واسط‌های مغز رایانه تلاش می‌کنند کلمات را از روی امواج مغزی بازسازی کنند تا این افراد بتوانند بدون صحبت کردن و تنها با فکر کردن به کلمات آن‌ها را به مخاطباشان برسانند. شناسایی کلمات از روی امواج مغزی می‌تواند با استفاد از روش‌های یادگیری ماشین و هوش مصنوعی انجام شود. در این پروژه، سیستمی هوشمند برای تشخیص گفتار متصور مربوط به ۴ کلمه و ۷ واج پیشنهاد شده است. این سیستم بر روی دیتاست کارا وان آموزش دیده‌است و استخراج ویژگی با استفاده از کراس-کوواریانس در حوزه زمان و فرکانس مورد آزمایش قرار گرفته‌است. در این پروژه نشان داده‌شده‌است که استفاده از کراس-کوواریانس روی حوزه فرکانس در زمان استخراج ویژگی نتایج بهتری نسبت به عدم استفاده از کراس-کوورانس و استفاده از سیگنال‌ها در حوزه زمانی دارد. در بخش کلاس‌بندی، چند معماری مختلف شبکه عصبی کانوولوشنی و استفاده از LSTM به همراه CNN آزمایش شده‌است. بهترین دقت به دست آمده ۴۳.۳۴ درصد در کلاس‌بندی هر ۱۱ کلاس ذکر شده است.

واژگان کلیدی: واسط مغز-رایانه، شبکه عصبی، شبکه عصبی کانوولوشنی، انتخاب ویژگی، دسته‌بندی، نوار مغزی، یادگیری ماشین

فهرست مطالب

فهرست شکل‌ها

فهرست جدول‌ها

فهرست کلمات اختصاری

Abbreviations Pages numbers

فصل اول: کلیّات

۱-۱ مقدمه

در سال‌های اخیر تمرکز بسیاری از پژوهشگران روی فهمیدن رمزگشایی کردن و بازشناختن گفتار متصور بوده‌است. صحبت کردن سازوکاری پیچیده است که نیازمند فعالیت بخش‌های مختلفی از مغز برای برنامه‌ریزی و کنترل دقیق عضلات مختلف بسیاری برای ادا کردن کلمات است.

واسط‌های مغز-رایانه (BCI) معمولا از تصور حرکت دادن یک نشانگر روی اسکرین استفاده می‌کنند. اما برخی تحقیقات تلاش کرده‌اند به بخش‌ها زبانی به طور مستقیم دست پیدا کنند. این واسط‌ها از روش‌های تهاجمی و غیرتهاجمی برای تشخیص فعالیت مغز استفاده می‌کنند. با وجود این که روش‌ها تهاجمی نسبت نویز به سیگنال کم‌تری دارند، به دلیل استفاده پیچیده آن‌ها تنها در موارد بسیار شدید کاربرد دارند. در نتیجه، تلاش بیش‌تر محققان روی روش‌های غیرتهاجمی است که بتوانند به صورت گسترده‌تر مورد استفاده قرار بگیرند.

پس از جمع‌آوری اطلاعات و سیگنال‌ها نوبت به پردازش سیگنال برای حذف نویز و مصنوعات اضافی می‌رسد، پس از آن باید به کم کردن داده‌ها با استخراج ویژگی و پیدا کردن مهم‌ترین بخش‌های سیگنال بپردازیم و در نهایت سیگنال‌ها را به کلمات مختلف دسته‌بندی کنیم.

در این پروژه تلاش شده‌است با آزمایش معماری‌های مختلف شبکه عصبی و استخراج ویژگی به دقت بالاتری نسبت به تحقیقات مشابه برسیم.

۱-۲ بیان مسئله

سیستم‌های بازشناخت گفتار متصور در سال‌های اخیر محبوبیت زیادی کسب کرده‌اند. این سیستم‌ها معمولا با هدف ثبت سیگنال‌های فعالیت مغزی از طریق روش‌های غیرتهاجمی یا تهاجمی، پردازش آن‌ها برای بالا بردن کیفیت سیگنال ، استخراج ویژگی‌های مهم برای تمرکز روی اطلاعات مهم سیگنال‌ها و در نهایت دسته‌بندی سیگنال‌ها برای دادن جواب به کاربر استفاده می‌شوند. این روند پردازشی نیازمند این است که سیستم به صورت بلادرنگ و روی وسایل قابل حمل انجام شود تا بتواند در زندگی هرروزه، برای کسانی که به آن نیاز دارند، مورد استفاده قرار بگیرد. کاربرد اصلی این سیستم‌ها کمک به افرادی است که به دلیل بیماری‌ها و اختلالات مختلف قدرت تکلم خود را از دست داده‌اند.

یکی از بخش‌های مهم ISR مرحله‌ی دریافت سیگنال‌های فعالیت مغزی است. روش‌های مختلفی برای ثبت کردن فعالیت مغزی برای استفاده از واسط‌های رایانه-مغز وجود دارد؛ مانند EEG، ECoG، MEG، PET و fMRI. رایج‌ترین شیوه‌ی مورد استفاده EEG است. بهترین مزیت EEG غیرتهاجمی و کم‌هزینه بودن آن است اما به دلیل این که سیگنال‌ها از روی جمجمه ثبت می‌شوند، لایه‌های زیادی بین الکترودها و مغز وجود دارد و به همین علت سیگنال‌های دریافت شده از این روش نویز بسیار زیادی دارند. دومین شیوه‌ی رایج برای دیکد کردن گفتار متصور ECoG است. داده‌ی جمع‌آوری شده از این روش به طور مستقیم از کورتکس مغز جمع‌آوری می‌شوند در نتیجه نویز کم‌تر و کیفیت بالاتری دارند. بنابراین این شیوه یک شیوه‌ی تهاجمی است به همین علت سختی‌های بیش‌تری نسبت به نوار مغز دارد.

محققانی که در حوزه گفتار متصور کار می‌کنند در سال‌های اخیر با ظهور دیتاست‌هایی که برای عموم قابل دسترسی هستند تمرکز خود را روی بالا بردن دقت دسته‌بندی کلمات از روی سیگنال‌های مغزی گذاشته‌اند. با توجه به این که بسیاری از پژوهش‌ها روی یک دیتاست انجام می‌شوند محققان می‌توانند کارهای خود را به سادگی با دیگران مقایسه کنند و نقاط ضعف و قوت پژوهش خود را بشناسند.

در این پروژه نیز تلاش شده‌است با استفاده از دیتاست کاراوان و معماری‌ها و ایده‌های مختلف، به دقت بالاتری برای دسته‌بندی کلمات و واج‌های این دیتاست دست یابیم.

۱-۳ کلیات روش پیشنهادی

با مطالعه پژوهش‌های پیشین ابتدا شیوه‌های مختلف استخراج ویژگی و کارایی آن‌ها شناسایی شده‌است. سپس تلاش شده این شیوه‌های مختلف استخراج ویژگی پیاده‌سازی شود و با استفاده از یک معماری شبکه عصبی واحد با یکدیگر مقایسه شوند. سپس با تغییر شبکه عصبی و آزمایش معماری‌های مختلف تلاش شده دقت بالاتر برود.

در بخش استخراج ویژگی ابتدا سیگنال‌های مربوط به بخش گفتار متصور از دیتاست کاراوان را به سگمنت‌ها ۲۵۰ میلی‌ثانیه‌ای تقسیم کردیم. ۵۰ درصد این سگمنت‌ها برای یادگیری و ۵۰ درصد دیگر برای تست استفاده شده است. در مرحله ابتدایی بدون هیچ پردازشی این سگمنت‌ها را به شبکه عصبی دادیم. پس از آن با استفاده از کراس-کوواریانس روی کانال‌های نوار مغز یک ماتریس به دست آوردیم و این ماتریس‌ها را به شبکه عصبی دادیم. سپس با ااستفاده از تبدیل فوریه سریع ابتدا سیگنال‌ها را به حوزه فرکانس برده و سپس دوباره روی آن‌ها کراس-کوورانس انجام دادیم.

پس از این مرحله روی معماری شبکه عصبی کانوولوشنی کار کردیم ابتدا معماری‌های مختلف شبکه عصبی کانوولوشنی ساده را آزمایش کردیم و سپس LSTM را نیز به آن اضافه کردیم.

۱-٤ ساختار پروژه

در این گزارش ابتدا به توضیح مفاهیم پایه و استفاده شده در پروژه می‌پردازیم. سپس دیتاست کاراوان که یک دیتاست دسترسی آزاد است را شرح می‌دهیم و به بررسی پژوهش‌های پیشین انجام شده در این زمینه می‌پردازیم. در نهایت کارهای انجام شده در این پروژه و نتایج به دست آمده را به طور کامل شرح می‌دهیم.

فصل دوم: مفاهیم پایه و کارهای مرتبط

۲-۱ مقدمه

گفتار متصور یا Imagined Speech به عمل فکر کردن به صحبت کردن و کلمات بدون حرکت دادن عضلات مربوط به صحبت کردن گفته می‌شود. در واقع گفتار متصور فکر کردن به حرف زدن بدون هیچ صدایی است. برای تشخیص گفتار متصور از واسط‌های مغز-رایانه (Brain-Computer Interface) استفاده می‌شود. واسط مغز-رایانه یا BCI مسیر ارتباطی‌ای میان فعالیت نورون‌های مغز و یک دستگاه خارجی ایجاد می‌کند. این دستگاه می‌تواند یک مانیتور، کامپیوتر یا عضوی رباتیک باشد. BCIها می‌توانند با هدف بازشناخت گفتار طراحی شوند. سیستم‌های بازشناخت گفتار (Speech Recognition Systems) به سیستم‌هایی گفته می‌شود که بتواند گفتار را از روی فعالیت مغزی تشخیص دهد و آن را دیکد کند و به متن تبدیل کند. برخی از این سیستم‌ها با صوت نیز کار می‌کنند. برای رسیدن به این هدف ابتدا باید به طریقی فعالیت‌های مغزی را ثبت کنیم.

۲-۱-۱ روش‌های جمع‌آوری سیگنال‌ها

چنانچه یک فرآیند پزشکی، از طریق ایجاد شکاف در پوست بوده و با تماس با موکوس، شکاف پوستی یا حفرات داخلی بدن (به جز حفراتی که از طریق منفذ طبیعی یا مصنوعی که از قبل وجود داشته قابل دسترسی هستند) انجام شود، تهاجمی و در غیر این صورت غیر تهاجمی خوانده می شود. مثالی از فرآیند تهاجمی، انجام تزریق و مثالی از فرآیند غیر تهاجمی، اندازه گیری فشار خون با استفاده از دستگاه های معمول فشار خون می باشد. جمع‌آوری اطلاعات فعالیت مغز نیز می‌تواند به روش تهاجمی یا غیر تهاجمی صورت بگیرد.

یکی از روش‌های غیرتهاجمی نوار مغز یا EEG است. نوار مغزی یا الکتروانسفالوگرافی (Electroencephalography) به عمل ثبت سیگنال توسط الکترودهای سطحی، تقویت سیگنال، حذف نویز، چاپ سیگنال و آنالیز آن می‌شود. به الکترودی که فعالیت امواج مغزی را ثبت می‌کند کانال نوار مغزی (EEG channel) گفته می‌شود. سیستم‌های نوار مغزی می‌توانند یک تا ۲۵۶ کانال داشته‌باشند. نحوه‌ی قرارگیری الکترودها روی سر از یک سیستم استاندارد که به آن سیستم ۲۰/۱۰ گفته می‌شود استفاده می‌کنند.

از روش‌های تهاجمی نیز می‌توان به ECoG اشاره کرد. ایکاگ یا الکتروکورتیکوگرافی نگاشت قشر مغزی است. این روشی مخرب است که در آن پس از جمجمه‌بری شبکه ای از الکترودها به درون قشر مغزی فرو برده می‌شوند. به دلیل این که این ناحیه مغز دارای سلول‌های حسی نیست، بنابراین شخص دردی احساس نمی‌کند و هوشیار است.

پس از دریافت سیگنل‌ها نوبت به پردازش و دسته‌بندی آن‌ها می‌رسد.

۲-۱-۲ CNN

از آنجا که استفاده از شبکه‌های عصبی تمام‌متصل (Fully connected) عمیق به قدرت محاسباتی (حافظه) بالایی نیاز دارد تا بتوان تعداد زیادی وزن و ضرب ماتریسی سنگین را مدیریت کرد، نوع جدیدی از شبکه‌های عصبی به‌نام شبکه‌ی عصبی کانولوشنی (Convolutional Neural Network) معرفی شده‌اند. در میان شبکه‌های عصبی، شبکه‌ی عصبی کانولوشنی یکی از بهترین‌ها برای حل مسائل حوزه‌ی بینایی ماشین (Computer Vision)، مانند شناسایی تصاویر (Image Detection)، طبقه‌بندی تصاویر (Image Classification)، تشخیص چهره (Face Recognition) و غیره، است.

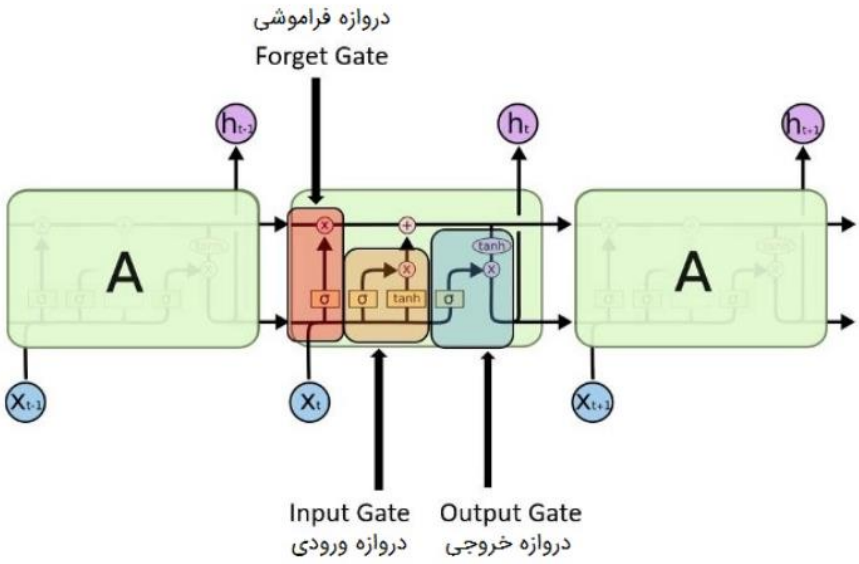
یک شبکه‌ی CNN از دو بخش کلی تشکیل شده است:

- استخراج ویژگی (Feature Extraction)
- طبقه‌بندی (Classification)

درواقع زمانی‌که یک عکس به یک شبکه‌ی CNN وارد می‌شود، ابتدا به مرحله‌ی استخراج ویژگی وارد می‌شود. در این مرحله هر عکس ورودی از چندین سری لایه‌ی کانولوشن (Convolution) و تابع فعال‌ساز ReLU و لایه‌ی pooling عبور می‌کند. سپس عکس‌های ورودی به طبقه‌بندی وارد می‌شوند؛ در این مرحله ابتدا مسطح‌سازی (Flattening) صورت می‌گیرد و سپس به یک لایه‌ی Fully Connected وارد می‌شوند و درنهایت یک تابع سافت مکس (Softmax) برای مسائل طبقه‌بندی چندکلاسه و یا تابع سیگموید (Sigmoid) برای مسائل طبقه‌بندی باینری روی آن اعمال می‌شود تا داده‌ها براساس مقادیر احتمالی میان صفر و یک طبقه‌بندی شوند.

۲-۱-۳ LSTM

شبکه‌های حافظه طولانی کوتاه مدت (Long Short-Term Memory) یک نسخه بهبود یافته از شبکه‌های عصبی بازگشتی هستند که باعث می‌شوند به خاطر سپردن داده‌های گذشته در حافظه، آسان تر شود. مشکل محوشوندگی تدریجی شبکه‌های عصبی بازگشتی در اینجا برطرف شده است. LSTM برای طبقه بندی، پردازش و پیش بینی سری‌های زمانی در حضور تأخیرهای زمانی با مدت نامشخص مناسب است. این شبکه، مدل را با استفاده از پس انتشار (back-propagation) آموزش می‌دهد. در یک شبکه LSTM، سه دروازه وجود دارد:



شکل ۱ معماری LSTM

- دروازه ورودی

تشخیص می‌دهد که از کدام مقدار ورودی باید برای بهبود حافظه استفاده شود. تابع سیگموئید (Sigmoid) تصمیم می‌گیرد که کدام مقادیر را از ۰و۱ عبور دهد. تابع tanh به مقادیر عبور کرده، بر اساس اهمیت آنها، وزنی در بازه ۱- تا ۱ می‌دهد.

- دروازه فراموشی

تشخیص می‌دهد چه جزئیاتی را باید از بلوک دور انداخت. این موضوع توسط تابع سیگموئید تصمیم گیری می‌شود. تابع سیگموئید، به حالت قبلی (ht-1) و ورودی محتوا (Xt) نگاه می‌کند و برای هر عدد در وضعیت سلول Ct-1، عددی بین ۰ (این را حذف کنید) و ۱ (این را نگه دارید) به عنوان خروجی برمی گرداند.

- دروازه خروجی

از ورودی و حافظه بلوک برای تصمیم گیری در مورد خروجی استفاده می‌شود. تابع سیگموئید تصمیم می‌گیرد که کدام مقادیر را از ۰و۱ عبور دهد. تابع tanh به مقادیر عبور کرده، بر اساس اهمیت آنها، وزنی در بازه ۱- تا ۱ می‌دهد و با خروجی تابع سیگموئید ضرب می‌شود.

۲-۱-۴ تبدیل فوریه سریع

تبدیل فوریه سریع (Fast Fourier Transform) یا FFT یکی از مهم‌ترین الگوریتم‌های مورد استفاده در پردازش سیگنال و آنالیز داده است. در واقع FFT یک الگوریتم است که برای محاسبه تبدیل فوریه گسسته (Discrete Fourier Transform) یا DFT و نیز معکوس آن (IDFT) مورد استفاده قرار می‌گیرد.

آنالیز فوریه می‌تواند یک سیگنال از حوزه اصلی، که معمولا زمان یا فضا است را به نمایشی در حوزه فرکانس و نیز بلعکس تبدیل کند. تبدیل فوریه گسسته معمولا از طریق تجزیه دنباله مقادیر، به عناصر با فرکانس‌های متفاوت محاسبه می‌شود. این تبدیل در بسیاری از رشته‌ها مفید است، اما مشکلی که وجود دارد این است که محاسبه مستقیم این تبدیل با استفاده از تعریف آن بسیار کند است و در عمل کاربردی ندارد. تبدیل فوریه سریع یا FFT روشی است که به وسیله آن می‌توان تبدیل فوریه گسسته را به سرعت محاسبه کرد. در واقع تبدیل فوریه سریع از طریق تجزیه ماتریس DFT به حاصلضرب ماتریس‌های تنک (Sparse) که در آن‌ها اکثر داریه‌های ماتریس صفر هستند، محاسبات را تسریع می‌بخشد.

افرادی به نام‌های کولی و توکی (Cooley and Tukey) توانستند الگوریتمی برای محاسبه تبدیل فوریه سریع یا Fast Fourier Transform به دست بیاورند. در این الگوریتم که مهم‌ترین الگوریتم تبدیل فوریه سریع است، به صورت بازگشتی

(Recursively) تبدیل فوریه گسسته را به مسایل کوچک‌تر می‌شکند و زمان مورد نیاز برای انجام محاسبات را به مقدار قابل توجهی کاهش می‌دهد.

با استفاده از تعریف تبدیل فوریه گسسته داریم:

$$
\begin{aligned}
X_k &= \sum_{n=0}^{N-1} x_n \cdot e^{-i\,2\pi\,k\,n\,/\,N} \\
&= \sum_{m=0}^{N/2-1} x_{2m} \cdot e^{-i\,2\pi\,k\,(2m)\,/\,N} + \sum_{m=0}^{N/2-1} x_{2m+1} \cdot e^{-i\,2\pi\,k\,(2m+1)\,/\,N} \\
&= \sum_{m=0}^{N/2-1} x_{2m} \cdot e^{-i\,2\pi\,k\,m\,/\,(N/2)} + e^{-i\,2\pi\,k\,/\,N} \sum_{m=0}^{N/2-1} x_{2m+1} \cdot e^{-i\,2\pi\,k\,m\,/\,(N/2)}
\end{aligned}
$$

در این حالت، تبدیل فوریه گسسته تکی را به دو عبارت تقسیم کردیم که هر کدام شباهت بسیار زیادی به عبارت تبدیل فوریه اصلی دارند و یکی بر روی اعداد فرد و دیگری بر روی اعداد زوج عمل می‌کنند. اما تا این قسمت هنوز هیچ توان محاسباتی را کاهش نداده‌ایم و هر عبارت از $\frac{N}{2} * N$ محاسبه تشکیل شده است که در مجموع N^2 محاسبه را شامل می‌شود.

تا زمانی که تبدیل فوریه کوچک‌تر دارای مقدار M زوج باشد، می‌توانیم این روش تقسیم و غلبه (Divide-and-Conquer) را به صورت تکراری انجام دهیم و هر بار هزینه محاسباتی را نصف کنیم. این روند را تا جایی ادامه می‌دهیم که آرایه به دست آمده آنقدر کوچک باشد که استفاده از این استراتژی دیگر تاثیری در بهبود محاسبات نداشته باشد.

۲-۱-۵ کوواریانس

یکی از شاخص‌های مهم وابستگی بین دو متغیر تصادفی (Random Variable) در آمار، کوواریانس (Covariance) است. این مفهوم به شکلی با پراکندگی و معیار واریانس (Variance) ارتباط دارد. البته واریانس مربوط به یک متغیر است در حالیکه محاسبه کوواریانس ارتباط بین دو متغیر را بوسیله پراکندگی‌هایشان نسبت به میانگین، نشان می‌دهد. هر چه

مقدار کوواریانس بین دو متغیر، بزرگتر باشد، میزان وابستگی بین آن‌ها بیشتر است و برعکس اگر میزان کوواریانس بین دو متغیر کم باشد، وابستگی خطی بین آن‌ها کم خواهد بود.

کراس-کوواریانس تعریف شده در این پروژه در واقع محاسبه کوواریانس بین هر کانال و تمام کانال‌های دیگر است.

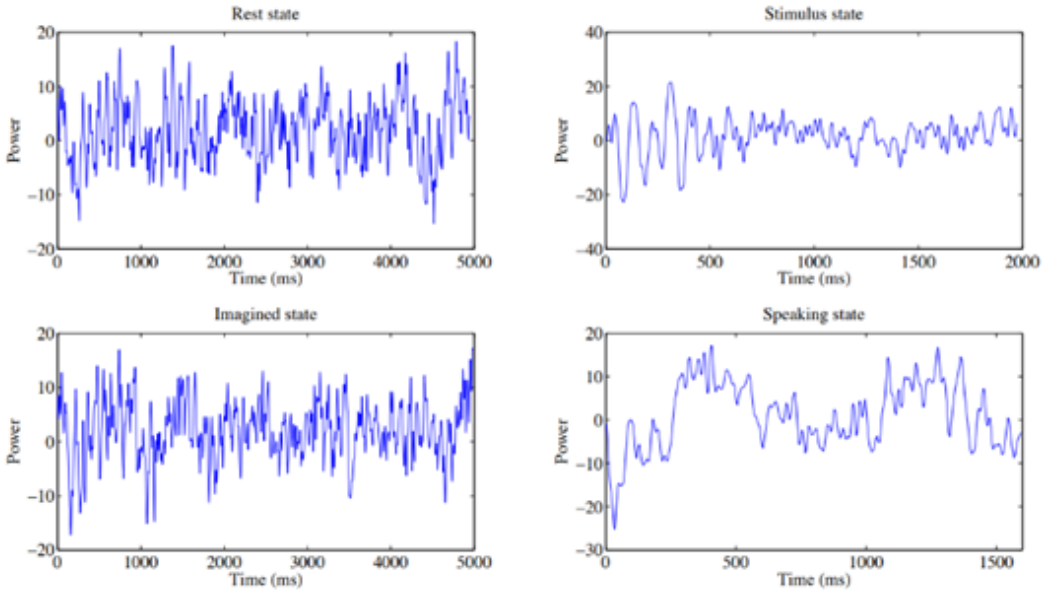
۲-۱-۶ دیتاست

در این پروژه از دیتابیس دسترسی آزاد ژائو و رودزیکز استفاده شده که در سال ۲۰۱۵ از موسسه توانبخشی تورنتو منتشر شد. در این دیتاست اطلاعات آوایی و حرکات صورت (با استفاده از کینکت) و سیگنال EEG افراد هنگام گفتار متصور وبیان کردن واج‌ها و کلمات مشخص شده، قرار دارد.

۱۴ شرکت‌کننده با میانگین سنی ۲۷ سال از دانشگاه تورنتو در این پروژه شرکت کرده‌اند. هیچ کدام از شرکت‌کنندگان سابقه بیماری اعصاب یا استفاده از مواد مخدر نداشتند. زبان مادری ده نفر از شرکت‌کنندگان انگلیسی بود و باقی نیز انگلیسی را در سطح پیشرفته صحبت می‌کردند.

هر کدام از شرکت‌کنندگان باید کارهای زیر به ترتیب انجام می‌دادند:

- حالت استراحت (۵ ثانیه): به شرکت‌کنندگان گفته می‌شد که ذهنشان را خالی کنند.
- حالت محرک: یک واج یا کلمه روی مانیتور به شرکت‌کننده نمایش داده می‌شد و صدای متناظر با آن از بلندگوها پخش می‌شد. به شرکت‌کنندگان گفته می‌شد که اداکننده واج‌هایشان را به حالتی دربیاورند که انگار می‌خواهند کلمه را به زبان بیاورند.
- حالت متصور (۵ ثانیه): شرکت‌کنندگان گفتن کلمه یا واج نمایش‌داده‌شده را تصور می‌شدند بدون این که آن را به زبان بیاورند.
- حالت صحبت: شرکت‌کنندگان کلمه یا واج نمایش‌داده‌شده را به زبان می‌آورند.

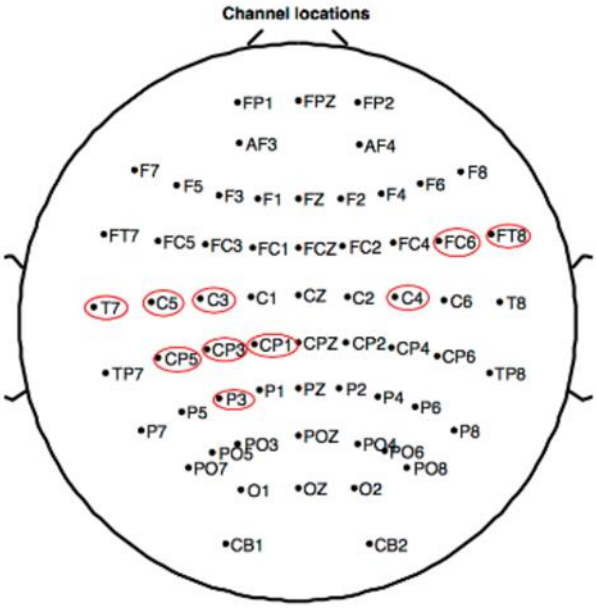


شکل ۲ امواج نوارمغز در حالت‌های مختلف دیتاست کاراوان

در این دیتاست ۷ واج /iy/، /uw/، /piy/، /tiy/، /diy/، /m/، /n/ و ۴ کلمه pat، pot، knew و gnaw انتخاب شده‌است. هر کدام از موارد ۱۲ بار به شرکت‌کنندگان نمایش داده می‌شد که یعنی برای هر شرکت‌کننده ۱۳۲ آزمایش وجود دارد. اول واج‌ها و سپس کلمات با ترتیب تصادفی به شرکت‌کنندگان نمایش داده می‌شدند.

در دیتاست استفاده شده روی دیتا پیش‌پردازش‌هایی نیز انجام شده‌بود. پیش‌پردازش با EEGLAB انجام شده و مصنوعات چشمی از سیگنال‌ها حذف شده. همچنین دیتا بین ۱ تا ۵۰ هرتز فیلتر شده و مقدار میانه از هر کانال کم شده‌است. همچنین فیلتر لاپلاسی روی هر کانال با استفاده از کانال‌های مجاور انجام شده.

هر سیگنال eeg به پنجره‌های مختلف با ۵۰ درصد هم‌پوشانی با پنجره‌های پیشین و پسین بخش‌بندی شده‌است و برای هر بخش آماره‌های مختلفی مانند ضریب چولگی، درجه اوج، انرژی و انروپی محاسبه شده. این محاسبات مجموعا ۶۵۸۳۵ ویژگی EEG (روی ۶۲ کانال) به ما می‌دهد. با محاسبه ضریب همبستگی پیرسون میان تمامی ویژگی‌ها ۱۰ کانال مغزی که بیش‌تری همبستگی را دارند مشخص شده‌اند که این همبستگی این ۱۰ کانال نشان‌دهنده‌ی تاثیر قشر حرکتی (Motor cortext) مغز روی تصمیم برای صحبت کردن است.



شکل ۳ کانال‌های دارای بیش‌ترین ضریب همبستگی

برای دیتاست تسک دسته‌بندی باینتری تعریف شده‌است.

- تشخیص مصوت یا صامت (C/V)
- تشخیص وجود یا عدم وجود مصوت دماغی (±Nasal)
- تشخیص وجود یا عدم وجود صامت دولبی (±Bilab)
- تشخیص وجود یا عدم وجود مصوت جلو-بالا (±/iy/)
- تشخیص وجود یا عدم وجود مصوت عقب-بالا (±/uw/)
- تشخیص حالت محرک یا صحبت (ST/SP)
- تشخیص حالت استراحت یا متصور (R/I)
- تشخیص حالت محرک یا متصور (ST/I)

آن‌ها از SVM برای دسته‌بندی این تسک‌ها استفاده کردند و بیش‌ترین دقت را در تشخیص وجود یا عدم وجود مصوت جلو-بالا به دست آوردند که ۷۹.۱۶ درصد بود. کم‌ترین دقت نیز متعلق به تشخیص مصوت و صامت با دقت ۱۸.۰۸ درصد بود.

۲-۲ تحلیل نقاط قوت و ضعف پژوهشی پیشین

در سال‌های اخیر، بازشناخت گفتار متصور از روی سیگنال‌های مغزی توجه تعداد زیادی از پژوهشگران را به خود جلب کرده‌است. رویکردهای مختلفی برای دستیابی به بهترین کارایی در سال‌های مختلف امتحان شده‌است.

اولین تلاش‌ها روی ساختن کلمه مورد نظر از روی حروف مختلف انجام شد. در این شیوه‌ها حروف با استفاده از کرسر روی مانیتور [1] یا دنبال کردن یک ماتریس نمادهای اسکی انجام می‌شد [2]. یکی از سیستم‌های بلادرنگ موفق که کلمات را حرف به حرف می‌ساخت توسط اوجوال چادهاری برای بیماری با ASL ساخته‌شد [3]. با توجه به کارایی بسیار پایین بخش حرکتی مغز در این بیمار، تنها را ارتباط با وی از طریق سیگنال‌های مغزی بود. این بیمار توانست بعد از چند روز آموزش در روز ۲۴۵ام جملات پیچیده بسیازد با میانگین یک حرف در دقیقه [3].

شیوه‌های این چنینی نتایج قابل‌توجهی داشتند و با استفاده بیش‌تر بیمار از سیستم نتایج بهتری نیز از خود نشان می‌دهند زیرا بیمار در طول زمان در استفاده از سیستم مهارت کسب می‌کند. اما این شیوه‌های ارتباطی غیرطبیعی و سخت است و نسبتا زمان زیادی برای ساختن یک کلمه نیاز دارند.

شیوه‌های دیگر روی ساختن کلمات مستقیما از روی دیکد کردن سیگنال‌های مغزی تمرکز کردند. این شیوه‌ها تصور می‌کنند هنگام تصور کلمات متلف مغز فعالیت‌های متفاوتی دارد که به نحوه‌ی تلفظ کلمات و وج‌ها مربوط است.

تیموتی پرویکس در تلاش برای یافتن نشانه‌های گفتار متصور از شیوه‌های تهاجمی ECoG استفاده کرد [4]. در پژوهش‌های انجام شد شباهت‌های بین سیگنال‌های مغزی هنگام گفتن کلمات و فکر کردن به آن‌ها پیدا شود. آن‌ها متوجه شدند که بازه فرکانسی ۸۰ تا ۱۵۰ هرتز هنگام گفتن و تصور کردن کلمات در بخش‌های حرکتی و حسگری زیاد شد در حالی که امواج بتا در این ناحیه‌ها کم شد. پژوهش دیگری که روی ECoG انجام شد توانست به نتایج مهمی هنگام دسته‌بندی پنج کلمه در یک سیستم مختص بیمار، دست پیدا کند. پژوهش‌گرا از ویژگی‌های زمانی گامای زیاد استفاده کردند. میانگین نتایج بدست آمده در این پژوهش ۵۸ درصد بود. در سال ۲۰۱۹ میگل انگریک توانست ترکیبی از سیگنال‌های ECoG و سیگنال‌های صحبت را استفاده کند که همزمان با ECoG گرفته شده بود. نتایج همبستگی‌ای بین سیگنال صحبت و خود صحبت پیدا کردند.

با این وجود هرچقدر هم که ECoG بتواند به نتایج قابل‌توجهی دست پیدا کند و کیفیت ثبت سیگنال بالایی داشته‌باشد باز هم روشی تهاجمی است و در نتیجه محدودیت‌های زیادی برای دریافت سیگنال‌ها و پذیرفته شدن توسط بیماران دارد. راه حل جایگزین برای این شیوه‌ی تهاجمی استفاده از fMRI، MEG و EEG است. تمام این روش‌ها نیز مشکلات خود را دارند. fMRI و MEG هزینه‌بر هستند و قابلیت حمل و نقل ندارند. در نهایت EEG بهترین شیوه برای ثبت فعلیت مغزی باقی می‌ماند با وجود این که نسبت نویز به سیگنال زیادی دارد.

در سال‌های اخیر تحقیقات زیادی روی استفاده از EEG برای تشخیص گفتار متصور انجام شده‌است. این یکی از این پژوهش‌ها [5]محققان ۶ کلمه مختلف را بررسی کردند که از ۱۵ شرکت‌کننده گرفته شده‌بود. در بخش استخراج ویژگی سگنال‌ها با استفاده از موج مادر db4 به ۸ مرحله تجزیه شدند که هر کدام معرف آلفا، بتا، گاما، و تتا بود به علاوه سه محدوده دیگر و ویژگی‌هایی مانند انحراف معیار و انرژی نسبی موجک‌ها محاسبه شدند. این ویژگیها به یک جنگل تصادفی (RF) و SVM داده شدند. نتایج برای هر دو شیوه‌ی کلاس‌بندی بالای دسته‌بندی شانسی بود که به طور میانگین ۲۵.۲۶ درصد برای جنگل تصادفی در بین ۱۵ شرکت‌کننده و ۲۸.۶۱ درصد برای SVM بود.

یکی از دستاوردهای مهم این حوزه زمانی بدست آمد که دیتاست‌های قابل دسترس برای همگان به وجود آمدند. [6] [7] این دیتاست‌ها به محققان کمک کردند که بدون نیاز به انجام عملیات سخت و پیچیده جمع‌اوری اطلاعات مغزی روی زمینه گفتار متصور تحقیق کنند و نتایج خود را دیگر پژوهش‌های انجام شده روی همان دیتاست مقایسه کنند.

با استفاده از دیتاست کارا وان، پاناچاکل و همکاران [8] توانستند سیستم مختص بیماری بر اساس ویژگی‌های آماری انحراف معیار، چولگی، گشتاور سوم و چیزهایی از این دست بر روی سیگنال‌ها پس از تجزیه سیگنال‌ها با استفاده از موجک مادر db4 به ۷ مرحله به دست بیاورند. پس از آن، سیگنال‌ها به یک شبکه یادگیری عمیق با دو لایه ۴۰ نورونی داده شدند. نتایج میانگین دقت ۵۷.۱۵ رای برای همه شرکت‌کنندگان نشان داد که نسبت به نتایج به دست آمده بدون یادگیری عمیق بالاتر است. با استفاده از معماری پیچیده‌تری از یادگیری عمیق، محققان دانشگاه بریتیش کلمبیا [9] توانستند سیستمی کلی برای دسته‌بندی باینری گفتار متصور طراحی کنند که این سیستم توانست به دقت ۸۵.۲۳ درصد برای تشخیص واج‌ها صامت یا مصوت دست پیدا کند. این نتایج با استفاده از ماتریس کوواریانس بین کانال‌های EEG و شبکه عصبی CNN و LSTM و یک اتوانکدر (DAE) به دست آمد.

در ۲۰۲۱ دیتاست بزرگ‌تری در مسکو جمع‌آوری شد که در آن ۲۷۰ شرکت‌کننده سالم وجود داشتند و اطلاعات گفتار متصور ۸ کلمه روسی از این شرکت‌کنندگان جمعاور شده بود. یک سیستم مختص بیمار روی این دیتاست توانست به دقت ۸۵.۴ درصد برای دسته‌بندی هر ۸ کلمه و ۸۷.۹ درصد برای دسته‌بندی باینری دست پیدا کند. این نتایج با استفاده از معماری شبکه عصبی عمیق ResNet18 در ترکیب با دو لایه GRU به دست آمد. در نهایت محققان این تحقیق ادعا کردند که سیگنال‌های شرکت‌کنندگان مختلف به شکل قابل توجهی با یکدیگر متفاوت است و دستیابی به دقت بالاتری با ساخت سیستم‌ها مختص بیمار بیش‌تر ممکن است به دست بیاید.

معماری‌های یادگیری عمیق نتایج بهتری در دسته‌بندی گفتار متصور نشان داده‌اند و در طی سال‌ها محبوبیت بسیار زیادی کسب کرده‌اند. اخیرا مطالعات EEG روی استفاده از LSTM به دلیل برتری‌ای که برای سری‌های زمانی دارد، متمرکز شده‌اند. LSTM نتایج قابل توجهی روی پیش‌بینی حمله صرع [10] و بازشناخت گفتار متصور نشان داده‌است [11].

۲-۳ جمع‌بندی

در بخش مقدمه مفاهیمی مانند واسط رایانه-مغز و گفتار متصور به طور کامل توضحی داده‌شد. همچنین موارد مربوط هوش مصنوعی و یادگیری عمیق استفاده در این پروژه کاملا شرح داده‌شدند. سپس به توضیح دیتاست و ویژگی‌های آن پرداختیم. پس از آن نیز پژوهش‌هایی که پیش از این در حوزه تشخیص گفتار متصور انجام شده‌بود را شرح دادیم و به طور ویژه پژوهش‌های انجام شده روی دیتاست مذکور را بررسی کردیم. در فصل بعد به نحوه پیاده‌سازی روش پیشنهادی و نتایج به دست آمده می‌پردازیم.

فصل سوم: روش پیشنهادی و نتیجه‌گیری

۳-۱ مقدمه

با توجه به مسئله‌ی مطرح شده و پژوهش‌های پیشین انجام شده روی این موضوع، در این پروژه تلاش شده که روش‌های تازه‌ای برای استخراج ویژگی و کلاس‌بندی اطلاعات نوار مغزی مورد آزمایش قرار بگیرد. پروژه روی دیتاست کاراوان انجام شده و هدف اصلی بالا بردن دقت در کلاس‌بندی تمام ۱۱ کلاس موجود اعم از واجها و کلمات است.

این کار از دو بخش کلی استخراج ویژگی و کلاس‌بندی تشکیل شده‌است. در بخش استخراج ویژگی از قسمت‌بندی کردن نوارمغزی مبوط به بخش گفتار متصور هر شرکت‌کننده شروع کردیم و شیوه‌های مختلفی مانند کراس‌کوواریانس و تبدیل فوریه سریع و حالت‌های مختلف این حالت‌ها را آزمایش کردیم. در بخش کلاس‌بندی نیز چند معماری مختلف از CNN و استفاده از LSTM را آزمایش کردیم.

برای پیاده‌سازی روش پیشنهادی از گوگل کولب، زبان پایتون و کتابخانه tensorflow برای بخش‌های لرنینگ استفاده شده‌است.

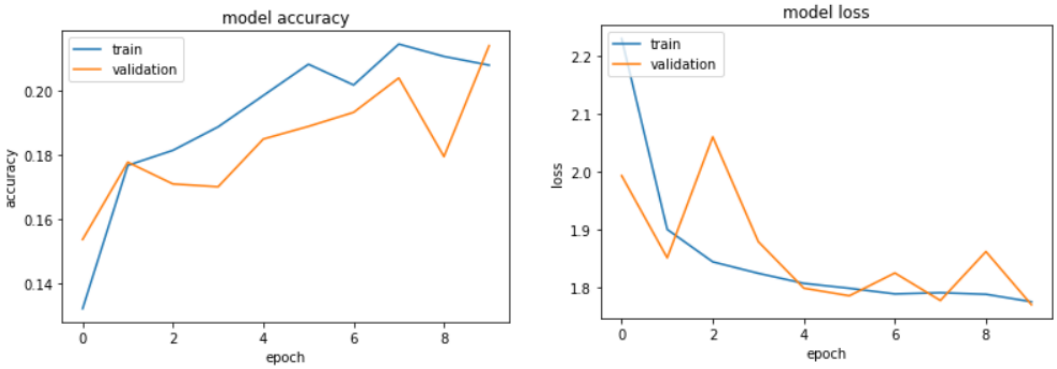
۳-۱-۱ سیگنال‌های حوزه زمان و CNN

سیگنال‌های داده زمانی هر کدام از سگمنت‌های جدا شده به ابعاد ۲۵۰ (اندازه هر سگمنت) و ۶۲ (تعداد کانال‌ها) به شبکه عصبی کانوولوشنی با ویژگی‌های زیر دادیم:

```
Train Shape: X (10712, 62, 250), Y (10712, 11)
Validation Shape: X (4592, 62, 250), Y (4592, 11)
Test Shape: X (15304, 62, 250), Y (15304, 11)
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_2 (Conv2D)           (None, 64, 250, 1)        35776

 conv2d_3 (Conv2D)           (None, 128, 250, 1)       73856

 flatten_1 (Flatten)         (None, 32000)             0

 dense_3 (Dense)             (None, 128)               4096128

 dense_4 (Dense)             (None, 64)                8256

 dense_5 (Dense)             (None, 11)                715

=================================================================
Total params: 4,214,731
Trainable params: 4,214,731
Non-trainable params: 0
```

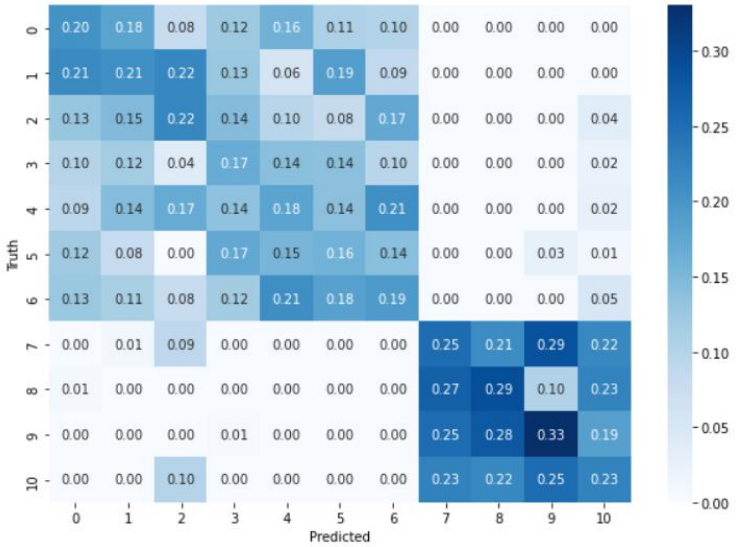
شکل ۴ معماری شبکه کانوولوشنی مرحله اول

عملیات یادگیری در ۱۰ دوره انجام شد و در نهایت به دقت ۲۰.۸۷ درصد روی داده‌های یادگیری و ۱۵.۶۶ درصد روی داده‌های تست رسید.



شکل ۵ مدل loss و accuracy مرحله اول

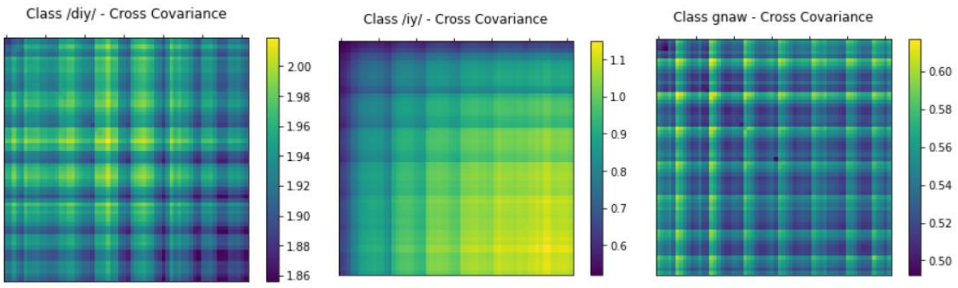
همانطور که از ماتریس کانفیوژن مشخص است این رویکرد در تشخیص کلمه یا واج به خوبی عمل می‌کند اما در کلاس‌بندی هر ۱۱ کلاس تنها به ۱۵.۶۶ درصد دقت توانسته است دست پیدا کند.



شکل ۶ ماتریس کانفیوژن مرحله اول

۳-۱-۲ کراس-کوواریانس در حوزه زمان به همراه CNN

در این مرحله روی هر کدام از سگمنت‌ها کراس-کوواریانس انجام شد. در نتیجه برای هر سگمنت، جمعه ۱۰۷۱۲ سگمنت مختلف، یک ماتریس ۶۲ در ۶۲ که ۶۲ تعداد کنال‌های نوار مغزی استفاده شده در این پروژه است به دست آوردیم.



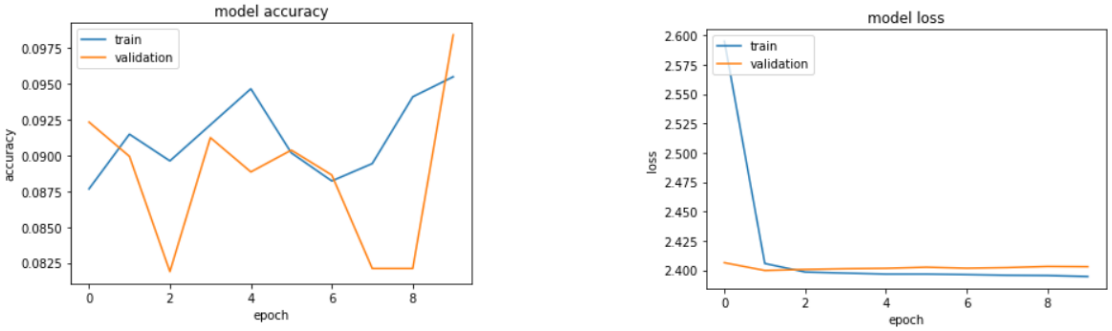
شکل ۷ ماتریس‌های کراس-کوواریانس حوزه زمانی

سپس ماتریس‌های به دست آمده را به شبکه CNN مشابه شبکه‌ای که در مرحله قبل استفاده کردیم دادیم.

```
Train Shape: X (10712, 62, 62), Y (10712, 11)
Validation Shape: X (4592, 62, 62), Y (4592, 11)
Test Shape: X (15304, 62, 62), Y (15304, 11)
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_2 (Conv2D)           (None, 62, 62, 64)        640

 conv2d_3 (Conv2D)           (None, 62, 62, 128)       73856

 flatten_1 (Flatten)         (None, 492032)            0

 dense_2 (Dense)             (None, 64)                31490112

 dense_3 (Dense)             (None, 11)                715

=================================================================
Total params: 31,565,323
Trainable params: 31,565,323
Non-trainable params: 0
```

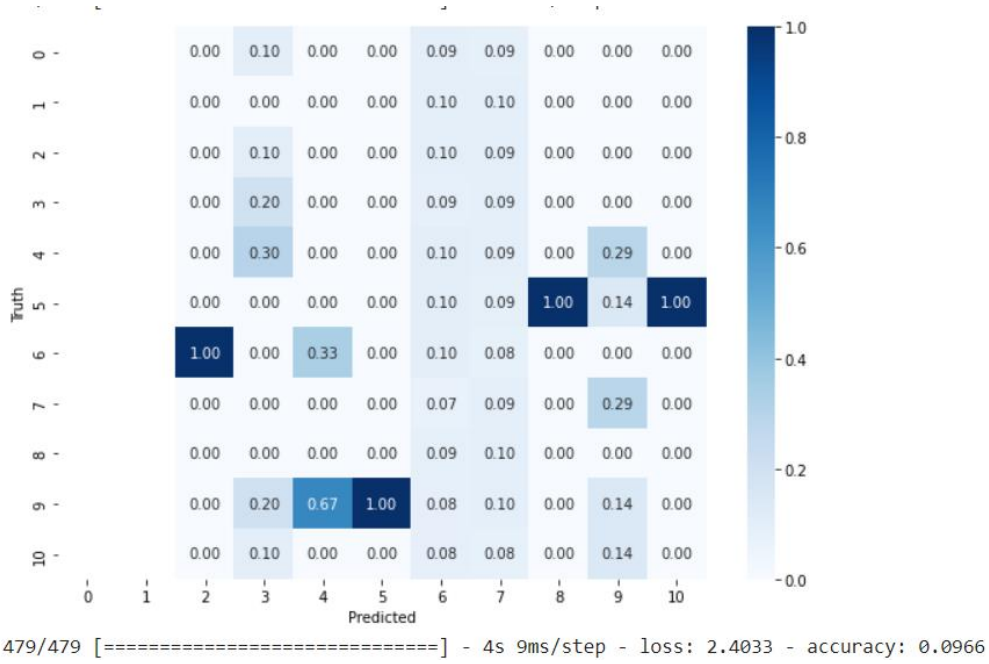
شکل ۸ معماری شبکه کانوولوشنی مرحله دوم

عملیات یادگیری در ۱۰ دوره انجام شد.



شکل ۹ مدل loss و accuracy مرحله دوم

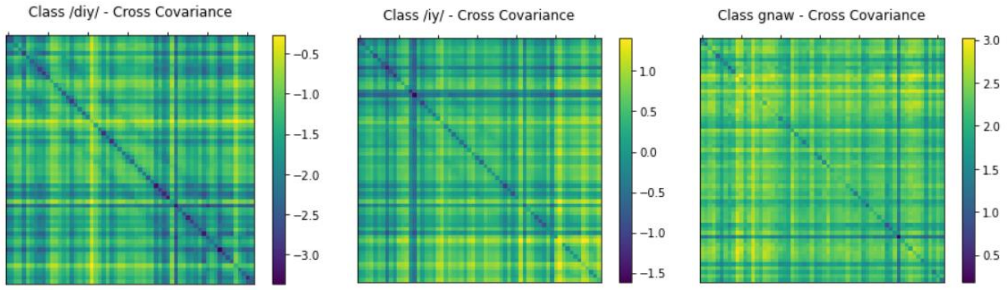
نتایج به دست آمده در این مرحله تقریبا در حد دسته‌بندی شانسی با دقت ۹.۶۶ درصد روی تست بود. این شیوه برای خلاف شیوه‌ی قبل در تشخیص کلمه و واج هم بسیار ضعیف عمل کرد و دقت کلی آن حدود دقت شانسی بود.



شکل ۱۰ ماتریس کانفیوژن مرحله دوم

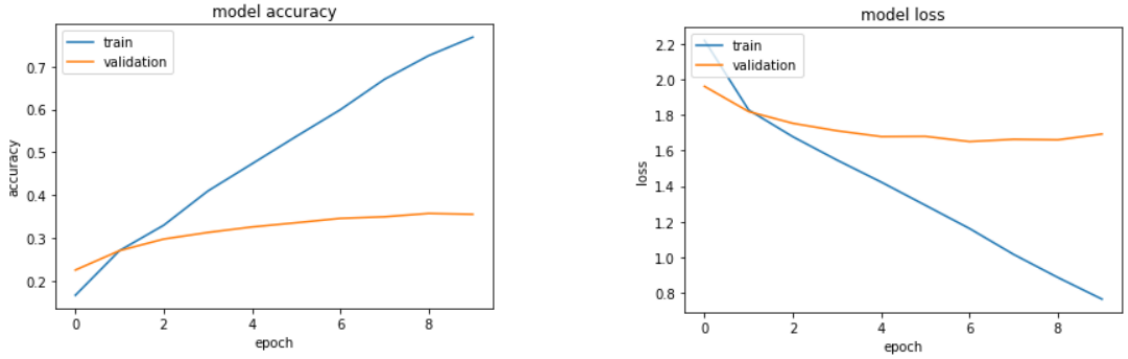
۳-۱-۳ کراس-کوواریانس در حوزه فرکانس به همراه CNN

این مرحله مشابه مرحله قبلی است با این تفاوت که پیش از انجام کراس-کوواریانس، سیگنال‌ها با استفاده از تبدیل فوریه سریع به حوزه فرکانس برده شده‌اند. نتیجه استخراج ویژگی این مرحله نیز مشابه قبل ماتریس‌های ۶۰ در ۶۲ است.



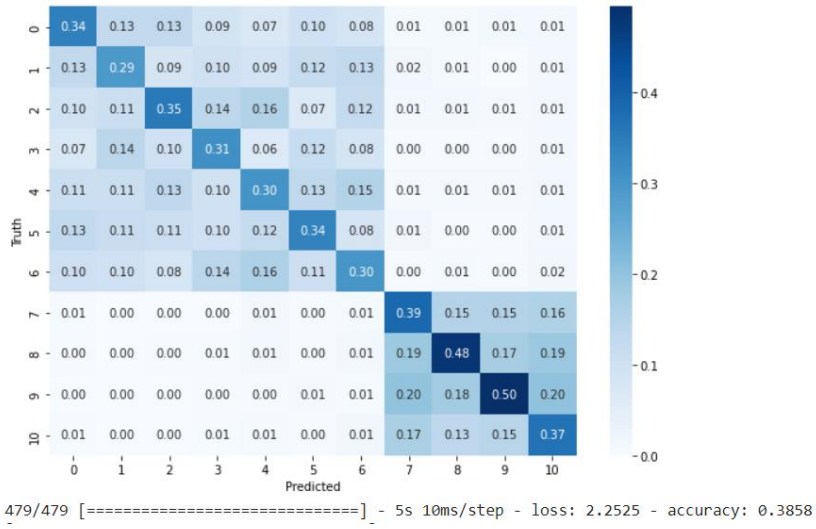
شکل ۱۱ ماتریس‌های کراس-کوواریانس حوزه فرکانس

سپس این ماتریس‌ها به همان شبکه کانوولوشنی مرحله قبل داده شدند.



شکل ۱۲ مدل loss و accuracy مرحله سوم

نتیجه‌ی به دست آمده در این مرحله نسبت به مرحله قبل به شکل قابل توجهی بهبود پیدا کرد و به ۳۸.۵۸ درصد رسید.



شکل ۱۳ ماتریس کانفیوژن مرحله سوم

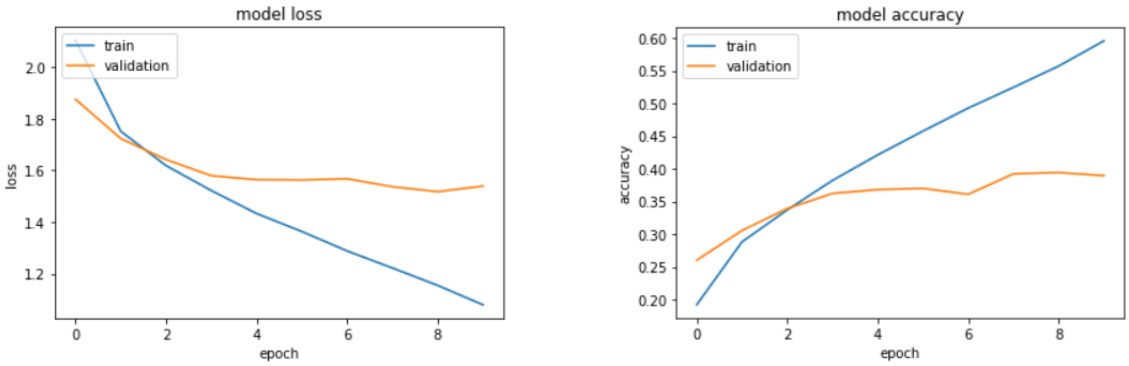
۴-۱-۳ کراس-کوواریانس در حوزه فرکانس به همراه CNN و LSTM

استخراج ویژگی در این مرحله مشابه مرحله قبلی است یعنی سیگنال‌ها ابتدا با استفاده از تبدیل فوریه سریع به حوزه فرکانس برده شده‌اند و سپس روی آن‌ها عملیت کراس-کوواریانس انجام شده‌است. با این تفاوت که در این مرحله در بخش کلاس‌بندی، شبکه عصبی LSTM نیز دارد.

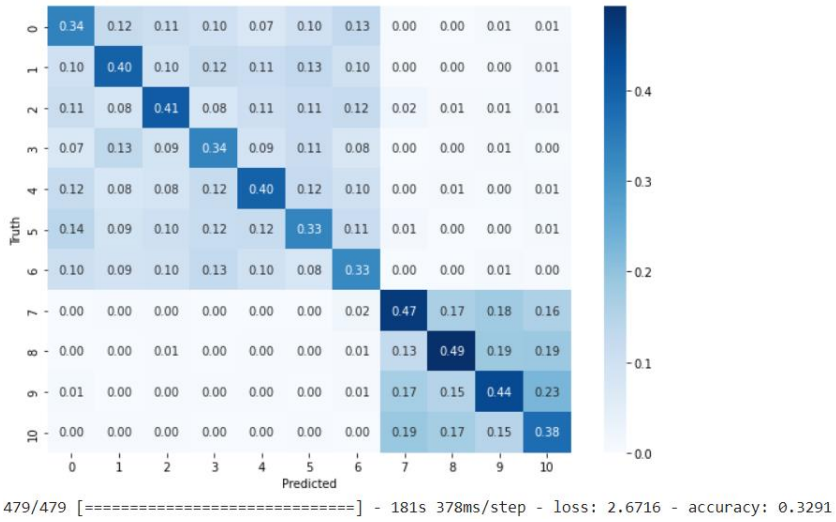
```
Train Shape: X (10712, 4, 62, 62), Y (10712, 11)
Validation Shape: X (4592, 4, 62, 62), Y (4592, 11)
Test Shape: X (15304, 4, 62, 62), Y (15304, 11)
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv_lstm2d (ConvLSTM2D)    (None, 4, 62, 62, 64)     150016

 conv_lstm2d_1 (ConvLSTM2D)  (None, 4, 62, 62, 128)    885248

 conv_lstm2d_2 (ConvLSTM2D)  (None, 4, 62, 62, 64)     442624

 flatten (Flatten)           (None, 984064)            0

 dense (Dense)               (None, 64)                62980160

 dense_1 (Dense)             (None, 11)                715

=================================================================
Total params: 64,458,763
Trainable params: 64,458,763
Non-trainable params: 0
```

شکل ۱۴ معماری شبکه عصبی همراه با LSTM

نتیجه:



شکل ۱۵ مدل loss و accuracy مرحله چهارم



شکل ۱۶ ماتریس کانفیوژن مرحله چهارم

۳-۲ روش ارزیابی

برای بررسی و مقایسه شیوه‌ها و ومعماری‌های مختلف استفاده شده در این پروژه از چند معیار کارایی مختلف استفاده شده‌است. ماتریس کانفیوژن یکی از معیارهای محبوب برای مسائل کلاس‌بندی است. در بخش قبلی برای هر کدام از بخش‌ها ماتریس کانفیوژن آورده شده‌است. ماتریس مذکور به شیوه زیر به دست می‌آید:

	مقدار پیش‌بینی‌شده					
	دسته	c_1	c_2	...	c_N	مجموع
مقدار واقعی	c_1	$o = c_1$	$o = c_2$	...	$o = c_N$	$\sum_j^N C_{c1j}$
	c_2	$o = c_1$	$o = c_2$	...	$o = c_N$	$\sum_j^N C_{c2j}$
	...	...	...	...	...	...
	c_N	$o = c_1$	$o = c_2$	...	$o = c_N$	$\sum_j^N C_{cNj}$
	مجموع	$\sum_j^N C_{jc1}$	$\sum_j^N C_{jc2}$	...	$\sum_j^N C_{jcN}$	$\sum_k^N C_{kk}$

جدول ۱ ماتریس کانفیوژن

دقت کلی نیز به شیوه‌ی زیر محاسبه می‌شود:

$$Accuracy = \frac{1}{N}\sum_{k=1}^{Nc} C_{kk}$$

که در آن N تعداد وکتورهای خروجی و Nc تعداد کلاس‌های دسته‌بندی است. C_{kk} نیز عناصر قطر اصلی ماتریس کانفیوژن است.

دقت متعادل (balanced accuracy) نیز برای زمانی که ورودی سیستم نامتعادل است استفاده می‌شود و فرمول آن به صورت زیر است.

$$Balanced\ Accuracy = \frac{1}{Nc}\sum_{k=1}^{Nc} \frac{C_{kk}}{t_k}$$

t_k در این فرمول تعداد دفعاتی است که کلاس k در داده‌ی ورودی ظاهر شده است و به شیوه‌ی زیر محاسبه می‌شود:

$$t_k = \sum_{j=1}^{Nc} C_{jk}$$

ضریب کاپا نیز معمولا برای محاسبه‌ی درجه‌ی موافقت مشاهدات مستقل مختلف استفاده می‌شود و فرمول آن به شیوه‌ی زیر است:

$$Kappa = \frac{C\ x\ N - \sum_k^{Nc} P_k \times t_k}{N^2 - \sum_k^{Nc} P_k \times t_k}$$

که در آن C تعداد کل پیش‌بینی‌های صحیح و p_k تعداد دفعاتی است که کلاس k پیش‌بینی شده‌ست.

معیار پوشش (Recall) نیز برای هر کدام از شیوه‌های محاسبه شده‌است که نحوه‌ی محاسبه آن به صورت زیر است:

$$Recall = \frac{C}{c + \sum_k^{Nc} fn_k}$$

که fn_k در آن تعداد دفعاتی است کلاس k به عنوان کلاس دیگری پیش‌بینی شده‌است.

۳-۳ نتایج

نتایج شیوه‌های مختلف با توجه به معیارهای توضیح داده شده در جدول زیر آورده شده است.

	accuracy	Balanced accuracy	recall	kappa
سیگنال زمانی و CNN	۱۵.۶۶	۱۵.۸۲	۱۵.۸۲	۰.۰۷
کراس-کوواریانس سیگنال زمانی و CNN	۹.۶۱	۹.۶۰	۹.۶۰	۰.۰۵
کراس-کوواریانس سیگنال فرکانسی و CNN	۴۳.۳۴	۴۳.۳۵	۴۳.۳۵	۳۷.۶۸
کراس-کوواریانس سیگنال زمانی و CNN و LSTM	۳۲.۹۱			

جدول ۲ نتایج به دست آمده در آزمایش‌های مختلف

۳-٤ جمع‌بندی

به طور کلی می‌توان گفت بردن سیگنال‌ها به حوزه فرکانس و استفاده از کراس-کوواریانس می‌تواند دقت سیستم را بالا ببرد. در آزمایش‌های انجام شده این درصد از ۱۵ روی سیگنال خام به ۴۳ درصد برای دسته‌بندی ۱۱ کلاس رسید. استفاده از LSTM دقت سیستم را بهتر نکرد اما با مقایسه مقادیر validation و accuracy می‌توان به این نتیجه رسید که استفاده از LSTM از اورفیت سیستم جلوگیری کرده و درصد validation و accuracy را به هم نزدیک کرده‌است.

منابع

[1] P. Kennedy, R. Bakay, M. Moore, K. Adams and J. Goldwaithe, "Direct control of a computer from the human central," *IEEE Trans. Rehab,* pp. 198-202, 2000.

[2] G. Jayabhavani and N. Rajaan, "Brain enabled mechanized speech synthesizer using Brain Mobile Interface," *Int. J. Eng. Technol,* vol. 5, pp. 333-339, 2013.

[3] U. Chaudhary, I. Vlachos, J. Zimmermann, A. Espinosa, A. Tonin, A. Jaramillo-Gonzalez, M. Khalili-Ardali, H. Topka, J. Lehmberg, G. Friehs and e. al., "Spelling interface using intracortical signals in a completely locked-in patient enabled via auditory neurofeedback training," *Nat. Commun,* vol. 13, p. 1236, 2022.

[4] T. Proix, J. Delgado Saa, A. Christen, S. Martin, B. Pasley, R. Knight, X. Tian, D. Poeppel, W. Doyle, O. Devinsky and e. al, "Imagined speech can be decoded from low- and cross-frequency intracranial EEG features," *Nat. Commun,* vol. 13, p. 48, 2022.

[5] Y. Varshney and A. Khan, "Imagined Speech Classification Using Six Phonetically Distributed Words," *Front. Signal Process,* vol. 1055, p. 2, 202.

[6] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," in *In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* South Brisbane, Australia, 2015.

[7] C. Nguyen, G. Karavas and P. Artemiadis, "Inferring imagined speech using EEG signals: A new approach using Riemannian manifold features," *J. Neural Eng,* vol. 016002, p. 15, 2018.

[8] J. Panachakel and A. Ramakrishnan, "Ananthapadmanabha, T.V. Decoding Imagined Speech using Wavelet Features and Deep Neural Networks," in *IEEE 16th India Council International Conference (INDICON)*, Rajkot, India, 2019.

[9] P. Saha, S. Fels and M. Abdul-Mageed, "Deep Learning the EEG Manifold for Phonological Categorization from Active Thoughts," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019.

[10] K. Tsiouris, V. Pezoulas, M. Zervakis, S. Konitsiotis, D. Koutsouris and D. Fotiadis, "A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals," *Comput. Biol. Med,* vol. 99, pp. 24-37, 2018.

[11] P. Agarwal and S. Kumar, "Electroencephalography-based imagined speech recognition using deep long short-term memory network," *ETRI J.*, vol. 44, pp. 672-685, 2022.

[12] S. Martin, P. Brunner, I. Iturrate, J. Millán, G. Schalk, R. Knight and B. Pasley, "Word pair classification during imagined speech using direct brain recordings," *Sci. Rep,* vol. 25803, p. 6, 2016.

[13] M. Angrick, C. Herff, E. Mugler, M. Tate, M. Slutzky, D. Krusienski and T. Schultz, "Speech synthesis from ECoG using densely connected 3D convolutional neural networks," *J. Neural Eng*, vol. 036019, p. 16, 2019.

[14] M. Angrick, M. Ottenhoff, L. Diener, D. Ivucic, G. Ivucic, S. Goulis, J. Saal, A. Colon, L. Wagner, D. Krusienski and e. al., "Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity," *Commun. Biol*, vol. 1055, p. 4, 2021.

[15] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard and e. al., "Tensorflow: A system for large-scale machine learning," in *12th Symposium on Operating Systems Design and Implementation*, Savannah, GA, USA, 2019.

واژه‌نامه

accuracy	دقت
back propagation	پس انتشار
balanced accuracy	دقت متعادل
brain computer interface	واسط مغز-رایانه
classification	کلاس‌بندی
computer vision	بینایی ماشین
convolutional neural network	شبکه عصبی کانوولوشنی
EEG	نوار مغز
face recognition	تشخیص چهره
feature extraction	استخراج ویژگی
flattening	مسطح‌سازی
fully connected	تمام‌متصل
image classification	طبقه‌بندی تصاویر
image detection	تشخیص تصویر
imagined speech	گفتار متصور
speech recognition system	سیستم بازشناخت گفتار
motor cortex	قشر حرکتی
recall	پوشش

پیوست

- لینک کولب:

https://colab.research.google.com/drive/1yaGFQQ8uyUbGtqxsSup6MWqfkMfPOLna?usp=sharing

- لینک گیت‌هاب:

https://github.com/erfanghobadian/imagined-speech-classification

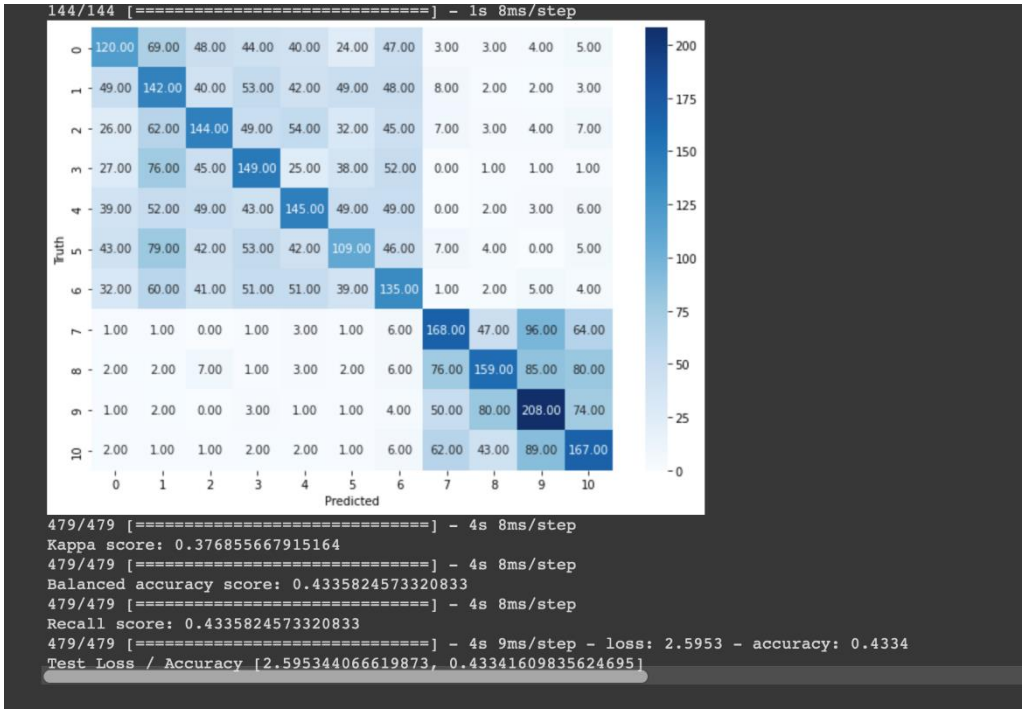
- کد کوواریانس، تبدیل فوریه سریع و نرمال‌سازی داده‌ها:

```
def std(x_train, x_test, x_val):
    mean = x_train.mean(axis=0)
    std_val = x_train.std(axis=0)
    x_train_std = (x_train - mean) / std_val
    x_test_std = (x_test - mean) / std_val
    x_val_std = (x_val - mean) / std_val
    return x_train_std, x_test_std, x_val_std


def fft(epoch):
    nfft = epoch.shape[1]
    freq = np.empty(1, dtype=int)
    freq[0] = int(nfft / 2)
    fft_res = np.fft.fft(epoch, n=nfft)
    fft_abs = np.abs(fft_res[:, :freq[0]])
    x = fft_abs * fft_abs
    x[x == 0] = 0.00001
    y = 20 * np.log(x)
    return y


def cov(epoch):
    c = np.cov(epoch)
    return c
```

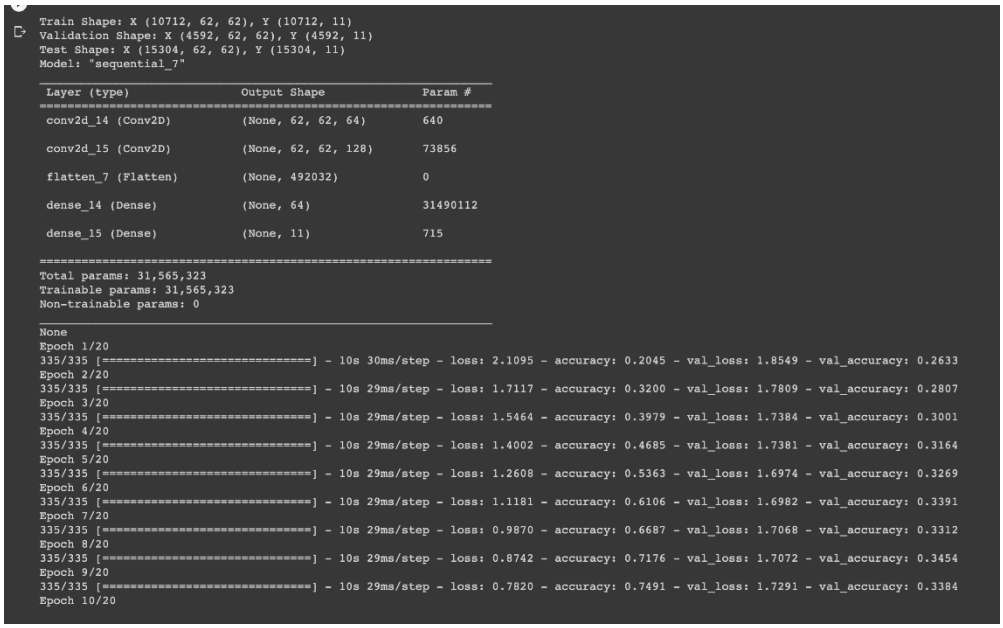
- کد بخش LSTM و شبکه عصبی کانوولوشنی

```
def create_lstm_model(self):
    model = tf.keras.Sequential([
        tf.keras.layers.ConvLSTM2D(
            64, (3, 3),
            strides=(1, 1), padding='same', activation='relu',
            recurrent_activation="sigmoid", data_format='channels_last',
            input_shape=self.input_shape, return_sequences=True
        ),
        tf.keras.layers.ConvLSTM2D(
            128, (3, 3),
            strides=(1, 1), padding='same', activation='relu',
            recurrent_activation="sigmoid", data_format='channels_last',
            return_sequences=True
        ),
        tf.keras.layers.ConvLSTM2D(
            64, (3, 3),
            strides=(1, 1), padding='same', activation='relu',
            recurrent_activation="sigmoid", data_format='channels_last',
            return_sequences=True
        ),
        tf.keras.layers.Flatten(),
        tf.keras.layers.Dense(64, activation="tanh"),
        tf.keras.layers.Dense(11, activation="softmax"),
    ])
    return model

def create_model(self):
    model = tf.keras.Sequential([
        tf.keras.layers.Conv2D(
            64, (3, 3), activation='relu', input_shape=self.input_shape,
            strides=(1, 1), padding='same',
        ),
        tf.keras.layers.Conv2D(
            128, (3, 3),
            activation='relu',
            strides=(1, 1), padding='same',
        ),
        tf.keras.layers.Flatten(),
        tf.keras.layers.Dense(64, activation="tanh"),
        tf.keras.layers.Dense(11, activation="softmax"),
    ])
    return model
```

- نتیجه‌ی آموزش روی LSTM و CNN – کراس-کووراینس حوزه فرکانس



- معماری شبکه عصبی کانوولوشنی به همراه LSTM



A Study on Neural Network Models for EEG Brain Signal for Imagined Words and Phenomes Classification

Abstract

Speech is a complex mechanism, which involves multiple brain areas in the process of production, planning, and controlling multiple muscles related to the utterance to create phenomes, words, and finally sentences. Speaking is one of the most important ways humans use to communicate. Some people are not able to speak due to some sickness and disorders. To facilitate these people's communication, Brain-Computer Interfaces try to recreate words from brain activities so that these people can communicate with other people without having to speak. Recognition of words from brain signals can be done using artificial intelligence and machine learning. In this project, an intelligent system is proposed for recognizing ۴words and ۷phenomes. The system has been trained on Kara One dataset and feature extraction is done using cross-covariance in the time and frequency domain. We showed that using cross-covariance in the frequency domain for feature extraction has better results than not using cross-covariance or using signals in the time domain. In the classification section, we examined multiple CNN architectures and LSTM. The best result accuracy in this project is ۴۳.۳۴for ۱۱classes.

Keywords: Brain-Computer Interface, neural network, convolutional neural network, feature selection, classification, EEG, machine learning

Shahid Beheshti University
Faculty of Computer Science and Engineering

A Study on Neural Network Models for EEG Brain Signal for Imagined Words and Phenomes Classification

By:
Erfan Ghobadian

A THESIS SUBMITTED
FOR THE DEGREE OF
BACHELOR OF SCIENCE

Supervisor
Dr. Yasser Shekofteh

February 2023