

Automatic sleep stage classification using a hybrid structure based on Frequency Multi-Head Attention mechanism and Convolutional Neural Network

Mehrpuya, Fatemeh¹ / Ebrahimi, Farideh^{2*}

¹ – MSc Student, Department of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran

² – Assistant Professor, Department of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran

ARTICLE INFO

DOI: <https://doi.org/10.22041/ijbme.2025.2069201.1994>

Received: 19-8-2025

Revised: 22-11-2025

Accepted: 18-12-2025

KEYWORDS

Attention Mechanism
Short-Term Fourier
Transform (STFT)
Convolutional Neural
Network (CNN)
Automatic Sleep Stage
Classification
EEG Signal

ABSTRACT

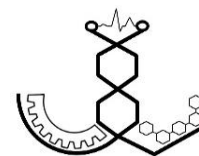
Approximately one-third of human life is spent in sleep, and sleep disorders can significantly affect quality of life and cognitive performance; therefore, accurate sleep stage detection is of particular importance. In this study, a lightweight parallel hybrid model is proposed for automatic sleep stage classification using EEG signals. In this model, features obtained by applying the multi-head attention mechanism (MHA) to the spectral content of EEG signals were combined with features extracted from one-dimensional convolutional layers (1D-CNN). Combining the spectral features improved by MHA with temporal features extracted from CNN resulted in a significant increase in classification accuracy. The proposed model achieved an overall accuracy of 81% in sleep stage classification on the Sleep_Edf_20 dataset, indicating its competitive performance over heavier architectures. To demonstrate the superiority of the proposed model, the one-dimensional convolutional neural network was combined with the multi-head attention mechanism in three different structures. The results showed that the proposed structure with a smaller number of parameters has greater accuracy in classifying sleep stages. In addition, to deal with the problem of data imbalance between different classes of sleep stages, a weighting method was used in the cost function, which provided a significant improvement in identifying undersampled stages, especially stage1 of sleep.

*Corresponding Author

Address: Department of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran

E-Mail: f.ebrahimi96@nit.ac.ir





طبقه بندی خودکار مراحل خواب توسط یک ساختار ترکیبی مبتنی بر مکانیزم توجه چند سر فرکانسی و شبکه عصبی پیچشی

مهرپویا، فاطمه^۱ / ابراهیمی خاله سری، فریده^{۲*}

^۱ - دانشجوی کارشناسی ارشد، گروه مهندسی پزشکی، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران
^۲ - استادیار، گروه مهندسی پزشکی، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران

مشخصات مقاله

شناسه‌ی دیجیتال: <https://doi.org/10.22041/ijbme.2025.2069201.1994>

پذیرش: ۱۴۰۴/۹/۲۷

بازنگری: ۱۴۰۴/۹/۱

ثبت در سامانه: ۱۴۰۴/۵/۲۸

چکیده

واژه‌های کلیدی

تقریباً یک سوم زندگی انسان در خواب سپری می‌شود و اختلالات خواب می‌توانند به طور قابل توجهی بر کیفیت زندگی و عملکرد شناختی تأثیر بگذارند؛ از این رو، تشخیص دقیق مراحل خواب از اهمیت ویژه‌ای برخوردار است. در این پژوهش، یک مدل ترکیبی موازی سبک‌وزن برای طبقه‌بندی خودکار مراحل خواب با استفاده از سیگنال‌های *EEG* ارائه شده است. در این مدل، ویژگی‌های حاصل از اعمال مکانیزم توجه چندسر (*MHA*) بر محتوای طیفی سیگنال‌های *EEG* با ویژگی‌های استخراج شده از لایه‌های پیچشی یک‌بعدی (*1D-CNN*) ادغام شدند. ترکیب ویژگی‌های طیفی بهبود یافته توسط *MHA* با ویژگی‌های زمانی استخراج شده از *CNN*، منجر به افزایش چشمگیر دقت طبقه‌بندی شد. مدل پیشنهادی بر روی مجموعه داده *Sleep_Edf_20* به دقت کلی ۸۱ درصد در طبقه‌بندی مراحل خواب دست یافت، که نشان‌دهنده عملکرد رقابتی آن نسبت به معماری‌های سنگین‌تر است. برای نشان دادن برتری مدل پیشنهادی، شبکه عصبی پیچشی یک بعدی، در سه ساختار مختلف با مکانیزم توجه چند سر ترکیب گردید. نتایج نشان داد که ساختار پیشنهادی با تعداد پارامتر کمتر، دقت بیشتری در طبقه بندی مراحل خواب دارد. بعلاوه برای مقابله با مشکل عدم توازن داده‌ها بین کلاس‌های مختلف مراحل خواب، از روش وزن دهی در تابع هزینه استفاده شد که بهبود محسوسی در شناسایی مراحل کم‌نمونه، به‌ویژه مرحله یک خواب ایجاد کرد.

مکانیزم توجه
تبدیل فوریه کوتاه مدت
شبکه عصبی پیچشی
طبقه بندی خودکار مراحل خواب
سیگنال EEG

*نویسنده‌ی مسئول

نشانی: گروه مهندسی پزشکی، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران

پست الکترونیک: f.ebrahimi96@nit.ac.ir



۱. مقدمه

خواب عبارت است از عدم توانایی ارادی در کنترل محیط فیزیکی و کاهش یا فقدان سطح هوشیاری، تعلیق نسبی ادراکات حسی و غیرفعال شدن تقریباً تمام عضلات ارادی در زمان استراحت، طوری که بدن و ذهن در حین خواب فعال است [۱]. داشتن خواب کافی و باکیفیت در هنگام شب می‌تواند به حفظ سلامت روانی، کیفیت زندگی، و بالا بودن سطح ایمنی کمک نماید. احساس ما در ساعت‌های بیداری تا حدودی به آنچه در هنگام خواب رخ می‌دهد وابسته است. در طول خواب، بدن همچنان به کار ادامه می‌دهد تا اینکه به حفظ سلامت جسمی و روانی کمک کند [۲]. اختلالات خواب به طور نزدیک با اختلالات عصبی و روانی مختلف مرتبط هستند [۳]. نشان داده شده است بین کاهش نسبت خواب عمیق در بیماران مبتلا به بیماری آلزایمر و شدت زوال عقل همبستگی وجود دارد [۴]. همچنین مشاهده شده است که بی‌خوابی منجر به دو برابر شدن خطر افسردگی در افراد، بدون علائم افسردگی می‌شود [۵]. از اینرو درمان به موقع و مؤثر بی‌خوابی می‌تواند در پیشگیری از ابتلا به افسردگی مؤثر باشد [۶]. گام اساسی در تشخیص و ارزیابی اختلالات خواب، طبقه‌بندی مراحل خواب (SSC¹) است که همچنین به عنوان مرحله بندی خواب یا امتیازدهی خواب نیز شناخته می‌شود [۷-۸]. در آزمایشگاه خواب، پلی سومنوگرافی^۲ (PSG) شامل پارامترهای فیزیولوژیکی مختلفی مانند فعالیت مغزی، حرکات چشم و فعالیت عضلانی است که در طول خواب با استفاده از چندین الکترود ثبت می‌شوند [۹]. برای ارزیابی خواب، PSG ابتدا به قطعات ۳۰ ثانیه‌ای (ایپاک^۳) تقسیم می‌شود و سپس هر ایپاک به صورت شهودی توسط کارشناسان بر اساس معیارهای خاصی در یک مرحله خواب طبقه بندی می‌شود. در قوانین R&K^۴ ثبت خواب شبانه به سه مرحله کلی تقسیم می‌شود [۱۰].

بیداری (W^۵)، حرکت سریع چشم (REM^۶) و غیر REM (NREM^۷) که NREM خود بر اساس عمق خواب به چهار زیرمرحله S1، S2، S3 و S4 تقسیم می‌شود. در استاندارد AASM^۸ ادغام مرحله ی S3 و S4 به N₃، مرحله بندی جدیدی به صورت W، N1، N2، N3 و REM ارائه شده است.

در روشهای یادگیری ماشین سنتی جهت طبقه بندی خودکار مراحل خواب، مرحله استخراج ویژگی بصورت دستی است. بنابراین استفاده از این روشها نیازمند دانش لازم در حوزه تحت پژوهش و تسلط کامل به روابط ریاضی است تا بتوان به درستی اطلاعات را استخراج نمود. از این رو مدل های مبتنی بر یادگیری عمیق در سال‌های اخیر به عنوان یک رویکرد محبوب برای طبقه‌بندی مراحل خواب شناخته شده است. در مقایسه با روشهای یادگیری ماشین سنتی، یادگیری عمیق قادر به استخراج خودکار ویژگی ها و یادگیری انتها به انتها^۹ است [۱۱]. در شبکه‌های عصبی عمیق تر، تعداد پارامترها اغلب زیاد است و این مساله ممکن است منجر به ایجاد مشکلات مربوط به ناپدید شدن گرادیان یا انفجار آن شود؛ همچنین غیریابستایی و سیگنال به نویز پایین سیگنال‌های EEG، می‌تواند بر نمایش ویژگی‌ها تأثیر منفی بگذارد. از اینرو در ادبیات تلاش شده است مدل‌هایی برای غلبه بر محدودیت‌ها ارائه گردد؛ در ادامه تعدادی از مطالعات انجام شده در ادبیات برای طبقه‌بندی خودکار مراحل خواب توسط یادگیری عمیق، مرور شده است. در یک مطالعه برای کاهش تعداد پارامترها و پیچیدگی محاسباتی، از تکنیک‌های تقویت داده و ساختار حافظه کوتاه بلند مدت (LSTM^{۱۰}) استفاده کرده‌اند و دقت بالای ۸۲٪ را در مجموعه داده‌های مختلف به دست آورده‌اند [۱۲]. در مطالعه دیگری از یک مدل ترکیبی شبکه عصبی پیچشی (CNN^{۱۱}) و LSTM برای طبقه‌بندی مراحل خواب در بیماران مشکوک به آپنه انسدادی خواب استفاده شد که به دقت ۸۳٫۷٪ در مجموعه داده Sleep-EDF رسیدند [۱۳]. مدل دیگری با استفاده از ترکیب CNN و LSTM دوطرفه، برای استخراج ویژگی‌ها و مدل‌سازی روابط زمانی ارائه شده است [۱۴]. برخی مدل‌ها برای کاهش تعداد پارامترها و هزینه محاسباتی، سبک‌تر طراحی شده‌اند. مدلی با استفاده از پیچش‌های نقطه‌ای و عمقی^{۱۲} و مازول توجه چندسره، ارائه شده است که عملکرد مناسبی بدست داده است [۱۵]. در یک پژوهش مدلی سبک‌وزن با هدف کاهش پیچیدگی محاسباتی ارائه گردیده است. در این مدل از شبکه‌ی پیچشی یک‌بعدی برای استخراج الگوهای زمانی سیگنال خواب و فشرده‌سازی ویژگی‌ها با لایه‌ی میانگین‌گیری سراسری (GAP^{۱۳}) انجام شده است. همچنین یک روش نرمال‌سازی تطبیقی (AdaBN^{۱۴}) به کار رفته است تا مدل در میان افراد

^۸ American Academy of Sleep Medicine

^۹ end-to-end learning

^{۱۰} Long Short Term Memory

^{۱۱} Convolutional Neural Network

^{۱۲} Point Wise & Depth Wise

^{۱۳} Global Average Pooling

^{۱۴} Adaptive Batch Normalization

^۱ Sleep stage classification

^۲ Polysomnography

^۳ epoch

^۴ Rechtschaffen&Kales

^۵ wakefulness

^۶ Rapid Eye Movement

^۷ Non-Rapid Eye Movement



افق‌های جدیدی در توسعه سیستم‌های قابل‌اطمینان و کاربردی، برای تحلیل سیگنال EEG گشوده است. به طور کلی ثبت خواب شبانه در منزل و اتاق خواب بیمار می‌تواند روی نتایج طبقه بندی بسیار مؤثر باشد. این مساله مستلزم طراحی مدلی با دقت تشخیص مناسب و همچنین با تعداد پارامتر کم است. در مطالعات قبلی نیز بر طراحی معماری‌های سبک با دقت بالا تاکید شده است. از این رو در این پژوهش، معماری جدیدی با عنوان SleepTFANet¹ برای طبقه بندی خودکار مراحل خواب پیشنهاد شده است. این مدل بر پایه یک ساختار موازی طراحی شده که به صورت هم‌زمان، ویژگی‌های زمانی و فرکانسی سیگنال EEG را استخراج می‌کند. در شاخه فرکانسی، با به کارگیری طیف انرژی مستخرج از تبدیل فوریه کوتاه مدت (STFT²) و مکانیزم توجه چندسر (MHA³) وابستگی میان باندهای طیفی مدل می‌شود. در مقابل، در شاخه زمانی با بهره‌گیری از پیش‌چس‌های چندمقیاسی، الگوهای زمانی کوتاه مدت و بلند مدت استخراج می‌گردند. در ادامه با ادغام اطلاعات مستخرج در دو مسیر، اطلاعات مکمل از هر دو دامنه زمان و فرکانس برای طبقه بندی مراحل خواب مورد استفاده قرار می‌گیرد و دقت طبقه بندی، به طور قابل توجهی افزایش می‌یابد. این طراحی موازی، در مقایسه با معماری‌های ترتیبی متداول، ضمن حفظ ویژگی‌های خام سیگنال، منجر به کاهش پیچیدگی محاسباتی و بهبود کارایی مدل شده است.

۲. ابزار و روش پیشنهادی

در این پژوهش، یک مدل مبتنی بر یادگیری عمیق برای طبقه بندی مراحل خواب از یک کانال سیگنال EEG طراحی و پیاده سازی شده است. این مدل شامل چندین لایه است که به ترتیب وظیفه استخراج ویژگی‌های محلی، شناسایی الگوهای مهم در داده‌ها، و انجام طبقه بندی نهایی را بر عهده دارند. ساختار کلی مدل در شکل ۱ آورده شده است و هر قسمت در ادامه تشریح شده است.

الف) لایه ورودی: داده‌های ورودی شامل سیگنال EEG خام هستند که به صورت برداری با طول مشخص به مدل وارد می‌شوند. این سیگنال‌ها پس از پیش پردازش اولیه (مانند نرمال سازی) برای ورود به لایه‌های بعدی آماده می‌شوند.

ب) بلوک پیچشی یک بعدی (Conv1D): در این بلوک، لایه‌های پیچشی یک بعدی، نرمال سازی دسته ای^۲ و لایه‌های تجمع بیشینه^۳ و تجمع میانگین^۴ برای کاهش بعد و فشرده

مختلف پایداری بیشتری داشته باشد. نتایج نشان داد این مدل با وجود پارامترهای کمتر، عملکردی قابل مقایسه با مدل‌های سنگین تر در طبقه بندی مراحل خواب ارائه می‌دهد [۱۶]. در مدل AttnSleep از شبکه ی عصبی پیچشی چند رزولوشنی و مکانیزم توجه چندسر بهره برده اند [۱۷]. در یک پژوهش با ترکیب CNN، LSTM و ماژول توجه، و طراحی تابع زیان جدید، مشکل عدم توازن کلاس‌ها را کاهش دادند [۱۸]. در پژوهش دیگری مدلی مبتنی بر گراف ارائه شده است در این تحقیق از ترکیب CNN گراف و CNN سه بعدی استفاده شده است و توانستند در مجموعه داده ISRUC-S3 به دقت ۸۳٪ برسند [۱۹]. در یک پژوهش برای مقابله با مشکل عدم توازن داده ها، از تابع هزینه آنتروپی متقاطع استفاده کرده اند و مشکل عدم توازن داده ها در کلاسهای مختلف را تعدیل و نتایج بهبود داده اند [۲۰]. از روش تکرار نمونه های اقلیت در تحقیق دیگری برای متوازن سازی داده ها در کلاسهای مختلف جهت طبقه بندی خواب استفاده شده است [۱۴]. از مدل ارائه شده ترکیبی CNN-BiLSTM که در آن همزمان از ساختار GAN و نویز سفید گوسی برای متوازن سازی داده ها استفاده کردند، نیز نتایج خوبی در طبقه بندی مراحل خواب حاصل شده است [۲۱]. در پژوهش دیگری یک چارچوب ترکیبی مبتنی بر CNN و CNN زمانی (TCNN) ارائه شده است. در این مدل، CNN برای استخراج ویژگی‌های محلی سیگنال EEG، و TCNN برای مدل سازی وابستگی‌های زمانی به کار رفته است. ترکیب این دو ساختار، امکان یادگیری مؤثر الگوهای زمان-مکان را فراهم کرده و به بهبود قابل توجه عملکرد طبقه بندی مراحل خواب منجر شده است [۲۲]. سرانجام در یک پژوهش، مدل سبک وزن MSA-CNN با ساختاری مبتنی بر ماژول چندمقیاسی و مکانیزم توجه زمانی معرفی گردیده است که با پارامترهای اندک، عملکرد قابل توجهی در طبقه بندی مراحل خواب ارائه کرده است [۲۳].

با توجه به مرور مطالعات پیشین، می‌توان نتیجه گرفت که محققان با استفاده از مدل‌های مبتنی بر یادگیری عمیق، به ویژه با ترکیب معماری‌های CNN و مکانیزم توجه، توانسته اند به عملکرد قابل قبولی در طبقه بندی خودکار مراحل خواب دست یابند. این مدلهای با حذف نیاز به استخراج دستی ویژگی‌ها، بهبود دقت، کاهش پیچیدگی محاسباتی و مقابله با عدم توازن داده‌ها، نسبت به روش‌های سنتی برتری یافته‌اند. همچنین، استفاده از ساختارهای سبک، شبکه‌های گراف و تکنیک‌های تقویت داده،

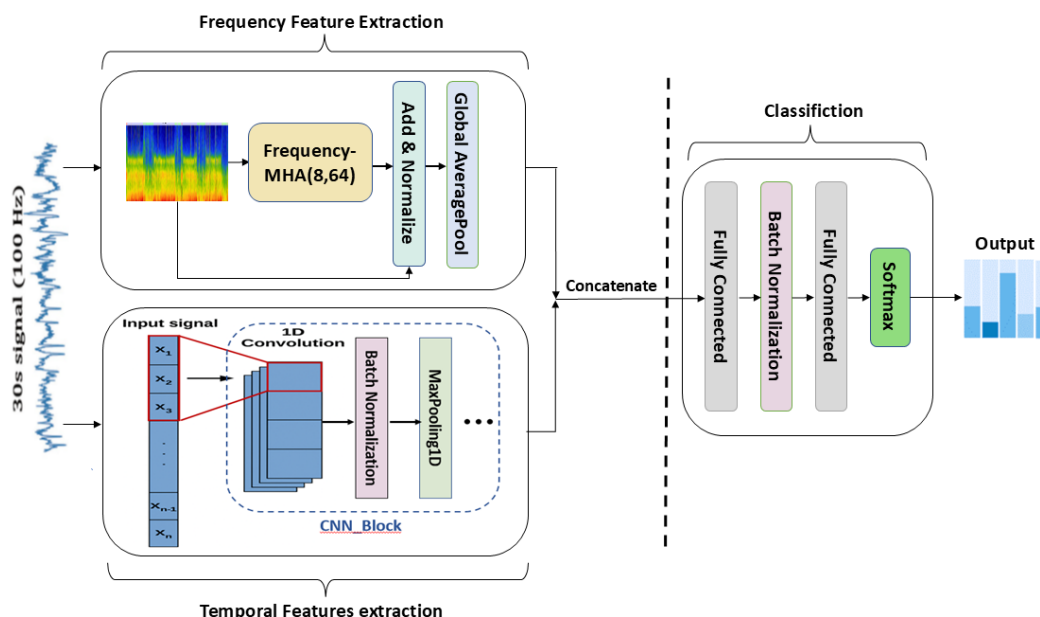
³ Multi-Head Attention

¹ Time-Frequency Attention for Sleep Staging

² Short-Time Fourier Transform

مربوط به مراحل خواب، و الگوهای خاص هر مرحله خواب استخراج گردند.
(پ) شاخه‌ی توجه فرکانسی

سازی وجود دارد. هر لایه پیچشی با تعداد متفاوتی فیلتر با اندازه‌های مختلف سبب می‌شوند الگوهای زمانی گوناگون در سیگنال EEG مانند تغییرات شکل موج، شروع و پایان الگوهای



شکل ۱: نمای کلی مدل پیشنهادی SleepTFANet برای طبقه‌بندی خودکار مراحل خواب

کمتری نسبت به سایر مراحل است، از تابع وزندهی در تابع خطا استفاده گردید. این روش باعث شد مدل، توجه بیشتری به کلاس‌های کم‌نمونه داشته باشد و در نتیجه دقت طبقه‌بندی در مراحل خواب با داده کمتر مانند مرحله ۱، افزایش یابد.

۲.۱ نمای کلی مدل

شکل ۱ چارچوب کلی مدل پیشنهادی را نشان می‌دهد. این مدل شامل سه بلوک اصلی است: (۱) استخراج ویژگی زمانی، (۲) استخراج ویژگی فرکانسی، (۳) طبقه‌بندی.

در این پژوهش، بخش‌های ۳۰ ثانیه‌ای خام از سیگنال EEG با نرخ نمونه‌برداری ۱۰۰ هرتز به عنوان ورودی مدل مورد استفاده قرار گرفتند. همان‌گونه که در شکل ۱ نشان داده شده است، سیگنال خام به صورت موازی برای استخراج ویژگی وارد دو شاخه‌ی مجزا می‌شود: شاخه پایین (شاخه زمانی) و شاخه بالا (شاخه فرکانسی). در شاخه پایین، سیگنال خام وارد لایه‌ی پیچشی یک‌بعدی (Conv1D) می‌شود تا ویژگی‌های زمانی سیگنال استخراج گردد. در مقابل، در شاخه بالا، ابتدا STFT بر سیگنال خام به منظور استخراج ویژگی‌های فرکانسی اعمال شده و خروجی حاصل وارد مکانیزم توجه چندسری می‌شود تا

در شاخه‌ی توجه فرکانسی معماری پیشنهادی، ابتدا نمایش زمان-فرکانس از سیگنال EEG با استفاده از STFT بدست می‌آید. در ادامه برای مدل‌سازی وابستگی میان باندهای فرکانسی، مکانیزم توجه چندسری در محور فرکانس طیف زمان-فرکانس (اسپکتروگرام) اعمال می‌گردد. این سازوکار به مدل امکان می‌دهد روابط معنادار میان باندهای فرکانسی مهم نظیر دلتا، تتا، آلفا و بتا شناسایی شده و ویژگی‌های طیفی به صورت مستقیم و بدون تخریب استخراج گردند. در مقایسه با معماری‌های ترتیبی، که در آن‌ها تبدیل STFT پس از لایه‌های پیچشی اعمال می‌شود و بخشی از ساختار طیفی اولیه ممکن است از بین برود، طراحی موازی پیشنهادی موجب حفظ کامل اطلاعات فرکانسی، بهبود تفسیرپذیری ویژگی‌ها و در عین حال کاهش پیچیدگی محاسباتی مدل شده است.

ت) طبقه بندی: ویژگی‌های زمانی و فرکانسی استخراج شده در دو مسیر موازی مدل پیشنهادی، ادغام و وارد یک ساختار شامل لایه‌های تماماً متصل^۱، Dropout و SoftMax می‌شوند تا بیداری و مراحل خواب تفکیک گردند.

ث) وزندهی تابع خطا: به منظور جبران عدم توازن داده‌ها در کلاس‌های مختلف، به‌ویژه مرحله ۱ خواب که دارای نمونه‌های

³ MaxPooling
⁴ AveragePooling

¹ Fully connected
² Batch Normalization Layer



هسته های بزرگ در لایه ی اول باعث می شود که ابعاد زمانی داده سریعتر کاهش پیدا کند و مدل سبک تر شود. بعلاوه تعداد فیلترهای کم (۳۲) در این لایه سبب می شود که ۳۲ ویژگی استخراج شود و مدل بجای تمرکز بر جزئیات، بر روی ویژگی های کلی تمرکز کند و موجب جلوگیری از یادگیری نویز و آرتیفکت خواهد شد. در نتیجه سرعت پردازش سریعتر می گردد. اگر تعداد کرنل ها برعکس یعنی از کوچک به بزرگ چیده شوند سرعت پردازش به شدت کاهش می یابد درحالیکه در دقت مدل، بهبودی ایجاد نمی گردد. درنهایت با گذاشتن فیلتر بزرگ در لایه آخر، مدل اطلاعات پیچیده با جزئیات بیشتری را استخراج می کند. گام های لایه های پیچشی به ترتیب از لایه ی اول ۵۰، ۶ و ۲ انتخاب شده است.

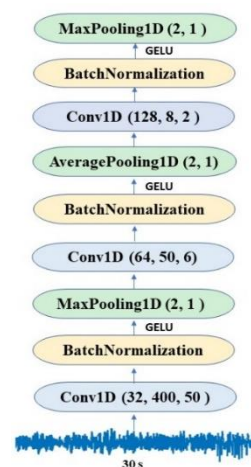
همانطور که در شکل ۲ نشان داده شده است، بعد از هر لایه پیچشی، از یک لایه نرمال سازی دسته ای یک بعدی^۱ استفاده شده است که با کاهش جابجایی کوواریانس داخلی، موجب می شود که مدل سریع تر همگرا شده و همچنین نرمال سازی به جلوگیری از ناپایداری های گرادینان مانند گرادینان ناپدید شونده یا منفجر شونده کمک می کند. سپس بعد از آن، از دولایه تجمیع بیشینه و یک لایه تجمیع میانگین استفاده شده است. وظیفه اصلی لایه های تجمیع، نمونه برداری فرعی (زیرمجموعه سازی) از نقشه های ویژگی است. این نقشه ها پس از عملیات پیچشی ایجاد می شوند. به عبارت دیگر، این لایه، نقشه های ویژگی با ابعاد بزرگ را به نقشه های کوچکتر کاهش می دهد و در عین حال، بیشتر اطلاعات غالب (یا ویژگیهای کلیدی) را در هر مرحله از فرآیند فشرده سازی حفظ می کند [۲۵]. علاوه بر این، عملیات نرمال سازی به سرعت بخشیدن به روند آموزش کمک می کند [۱۷].

در مدل پیشنهادی از تابع فعالساز $GELU^2$ استفاده شد. این تابع فعالسازی برای مدل های پیچیده، مناسب تر است و با اینکه سرعت پردازش را کمی پایین آورد اما دقت را بهبود می دهد. در مراحل انجام پژوهش، در ابتدا از تابع فعالساز $ReLU^3$ استفاده شد اما زمانی که مکانیزم توجه به مدل اضافه گردید دیده شد که تابع فعالساز $GELU$ انتخاب بهتری است زیرا برخلاف $ReLU$ ، در زمانهایی که ورودی منفی است، به طور کامل صفر نمی شود، بلکه بر اساس توزیع نرمال، مقداری نرم تر و غیر صفر در خروجی ایجاد می کند. این ویژگی باعث می شود $GELU$ عملکرد بهتری در مدل های عمیق، به ویژه در معماری های مبتنی بر توجه، داشته باشد.

وابستگی میان باندهای فرکانسی و الگوهای طیفی مهم، مدل سازی شود. در نهایت، ویژگی های استخراج شده از هر دو شاخه، پس از فشرده سازی توسط لایه ی GAP با یکدیگر ادغام شده و به بخش طبقه بندی نهایی ارسال می شوند. در ادامه جزئیات هر قسمت از مدل پیشنهادی توضیح داده شده است.

۲.۲ استخراج ویژگی زمانی

برای استخراج ویژگی های زمانی، یک مدل پیچشی یک بعدی طراحی شد که معماری آن در شکل ۲ نشان داده است. این مدل شامل سه لایه پیچشی با اندازه فیلتر های متفاوت برای استخراج اطلاعات است بدین صورت که اول، فیلتر بزرگ (اندازه ۴۰۰ نمونه) گام های زمانی را با پنجره های ۴ ثانیه ای طی می کند، و بنابراین اطلاعات موجود در سیگنال EEG با فرکانس نمونه برداری ۱۰۰ هرتز را تا ۰.۲۵ هرتز استخراج می کند. دوم، برای فیلتر متوسط (اندازه ۵۰ نمونه)، هر پنجره گام ۵۰ نمونه (۰.۵ ثانیه) را طی می کند، بنابراین قادر خواهد بود اطلاعات را تا ۲ هرتز استخراج کند. بدین ترتیب اطلاعات در باند فرکانسی پایین دلتا که در خواب عمیق غالب است استخراج می گردد و برای فیلتر کوچکتر (با اندازه ی ۸ نمونه)، هر پنجره گام ۸ نمونه (۰.۰۸ ثانیه) را طی می کند و قادر خواهد بود اطلاعات را تا ۱۲ هرتز استخراج کند که این باند فرکانسی مربوط به دوکهای خواب در مرحله ی ۲ خواب (باند فرکانسی ۱۲ تا ۱۴ هرتز) [۲۴] و امواج مربوط به خواب REM (باند فرکانسی ۸-۳۰ هرتز) هستند. به طور کلی با استخراج اطلاعات تا فرکانس ۱۲ هرتز امواج دلتا، تتا، آلفا به نحو مناسبی استخراج می گردند. تعداد فیلترها در لایه های مختلف بصورت معکوس با اندازه آنها مورد استفاده قرار گرفتند، به این صورت که به هسته ی بزرگتر، تعداد فیلتر کمتر اختصاص داده شده است و بالعکس. قرار دادن



شکل ۲. بلوک CNN برای استخراج ویژگی

³ Rectified Linear Unit

¹ 1D batch normalization

² Gaussian Error Linear Unit

۲.۲ استخراج ویژگی فرکانسی

جهت استخراج اطلاعات فرکانسی پس از اعمال STFT، طیف توان سیگنال در فرکانسهای مختلف استخراج می‌شود و طیف توان در هر فرکانس در زمانهای مختلف، به عنوان توکن در مکانیزم توجه مورد استفاده قرار می‌گیرند. در ادامه جزئیات مربوط به استخراج ویژگی‌های فرکانسی توضیح داده شده است.

۲.۳.۱ تبدیل فوریه کوتاه مدت

برای استخراج اسپکتروگرام از سیگنال EEG، ابتدا STFT بر سیگنال اعمال می‌شود. بدین منظور هر قسمت ۳۰ ثانیه‌ای از سیگنال به قسمتهای زمانی کوتاه برش شده و برای هر قسمت تبدیل فوریه به صورت زیر محاسبه می‌شود:

$$X[m, k] = \sum_{n=0}^{l-1} x[n + mH].w[n].e^{-j2\pi kn/N} \quad (1)$$

در رابطه فوق $x[n]$ قسمتهای به طول $l=200$ نمونه از سیگنال EEG هستند که با همپوشانی $H=100$ نمونه (۵۰٪ همپوشانی) مورد استفاده قرار گرفته اند و $w[n]$ تابع پنجره همینگ است. برای هر بخش از سیگنال، تبدیل فوریه با $N=256$ روی سیگنال اعمال می‌شود. پارامتر m نیز مشخص کننده نمونه های زمانی سیگنال است. این تقسیم بندی سبب می شود ویژگی های فرکانسی سیگنال یا نحوه تغییرات الگوهای فرکانسی سیگنال در طول زمان، مشخص شوند. بعد از اعمال STFT، اسپکتروگرام به صورت یک ماتریس که نمایش زمان-فرکانس انرژی سیگنال است، بدست می‌آید. اسپکتروگرام استخراج شده یک ماتریس دوبعدی با ابعاد (تعداد فریم ها $\times$ تعداد بین های فرکانسی) است که به عنوان ورودی بخش توجه فرکانسی مدل، مورد استفاده قرار می‌گیرد. تمرکز مدل روی توکن های فرکانسی، سبب می‌شود نه تنها ویژگی های کلی سیگنال، بلکه تغییرات محلی و لحظه ای در فرکانسهای مختلف نیز استخراج گردد؛ این ویژگی به ویژه برای سیگنالهای EEG که غیریستا هستند از اهمیت زیادی برخوردار است. [۲۶].

۲.۳.۲ مکانیزم توجه چند سر

در این پژوهش، برای مدل سازی وابستگی های بلندمدت میان باندهای فرکانسی و شناسایی روابط سراسری در ساختار طیفی سیگنال EEG، از مکانیزم توجه چندسر استفاده شد. این سازوکار با الهام از مدل ترنسفورمر [۲۷-۱۷]، قادر است با تخصیص وزن های یادگیرنده به بخش های مختلف ورودی، تمرکز مدل را بر باندهای موثرتر افزایش دهد. پس از اعمال STFT و محاسبه اسپکتروگرام، نمایش طیفی سیگنال EEG به صورت دنباله ای از بردارهای ویژگی در محور فرکانس به

عنوان توکن مورد استفاده قرار می‌گیرند. در این حالت، لایه ی توجه چندسر، روابط میان این باندهای فرکانسی را بر اساس شباهت میان بردارهای پرسش (Q) و کلید (K) محاسبه کرده و وزن های حاصل را بر بردارهای ارزش (V) اعمال می‌کند. فرمول کلی توجه در هر سر به صورت زیر تعریف می‌شود:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

در این رابطه پارامتر d_k نمایانگر ابعاد بردارهای پرسش و کلید در هر سر توجه و تعیین کننده ی ظرفیت مدل در نمایش ویژگی های وابسته به زمان است. توجه به محور فرکانس در مدل پیشنهادی، باعث می‌شود مدل بتواند وابستگی های پیچیده و تعامل میان باندهای دلتا، تتا، آلفا و بتا را به صورت داده محور بیاموزد و در نتیجه، ویژگی های فرکانسی غنی تر و قابل تفسیرتری استخراج کند. در این ساختار، سه بردار Query، Key و Value همگی از اعمال یک نگاشت مشترک (dence layer) روی ورودی فرکانسی به دست می‌آیند. این ساختار به عنوان مکانیزم خود توجه (Self-Attention) شناخته می‌شود و موجب کاهش تعداد پارامترها و ساده سازی معماری می‌شود، بدون آن که توان مدل در یادگیری وابستگی های فرکانسی کاهش یابد. ابعاد خروجی این بخش به صورت مستقیم از حاصل ضرب تعداد سرها و بُعد کلید تعیین می‌شود که مطابق با استاندارد معماری های مکانیزم توجه باید با بُعد کلی مدل تطابق داشته باشد. برای مثال، در حالتی که ۱۶ سر توجه و بُعد کلید برابر با ۸ انتخاب شده باشد، خروجی نهایی مکانیزم توجه دارای بُعد ۱۲۸ خواهد بود. در معماری حاضر، با افزایش بُعد کلید به مقدار ۶۴ (همراه با ۸ سر توجه)، ابعاد خروجی به ۵۱۲ افزایش یافته است. این افزایش ظرفیت محاسباتی، در نگاه اول ممکن است منجر به نگرانی از افزایش هزینه پردازش شود، اما نتایج تجربی نشان داده اند که این افزایش، بدون تحمیل بار محاسباتی سنگین قابل اجرا بوده و موجب بهبود عملکرد مدل در استخراج الگوهای پیچیده تر می‌شود. لایه ی مکانیزم توجه چندسر به صورت پیش فرض دارای لایه Dense است که ابعاد خروجی مکانیزم توجه را با ابعاد ورودی توکن ها تطبیق می‌دهد. به همین دلیل، حتی با تغییر ابعاد توکن ها و تعداد سرها، سازگاری ابعادی حفظ می‌شود.

۲.۳.۳ شاخه ی توجه فرکانسی

به منظور استخراج اطلاعات فرکانسی، ابتدا اسپکتروگرام سیگنال EEG در ایپاکهای ۳۰ ثانیه ای استخراج می‌شود که تغییرات انرژی در باندهای مختلف طی زمان با ابعاد



مطلوبی میان توانایی نمایش غیرخطی و کارایی محاسباتی فراهم می‌سازد. به منظور افزایش قابلیت تعمیم پذیری مدل و کاهش احتمال بیش‌برازش، یک لایه dropout با نرخ ۰.۵ اعمال شده است که به صورت تصادفی برخی از نورون‌ها را در طول آموزش غیرفعال می‌سازد. در پایان، طبقه‌بندی نهایی از طریق یک لایه کاملاً متصل همراه با تابع فعال‌ساز Softmax انجام می‌گیرد تا توزیع احتمالات به منظور تعیین مراحل خواب به دست آید.

۲.۵ تابع زیان و بهینه‌سازی

برای آموزش مدل پیشنهادی، از تابع زیان آنتروپی متقاطع دسته‌ای استفاده شد. این تابع برای مسائل طبقه‌بندی چندکلاسه مناسب است و به صورت زیر تعریف می‌شود:

$$L = -\frac{1}{M} \sum_{i=1}^M \sum_{k=1}^K y_{i,k} \log(p_{i,k}) \quad (3)$$

که در آن $y_{i,k}$ برچسب واقعی برای نمونه i ام در کلاس k ، $p_{i,k}$ احتمال پیش‌بینی شده توسط مدل برای نمونه i ام در کلاس k ، که در آن k تعداد کلاس‌ها و M تعداد کل نمونه‌ها است. این تابع زیان، تفاوت بین توزیع پیش‌بینی‌شده و توزیع واقعی را اندازه‌گیری می‌کند و به مدل کمک می‌کند تا با کمینه‌سازی این تفاوت، دقت پیش‌بینی‌ها را افزایش دهد. همچنین در این مدل برای حل مشکل عدم توازن تعداد داده‌ها در کلاسهای مختلف، از یک تابع وزن دهی استفاده شده است که با توجه به

تعداد نمونه‌ها در هر کلاس یک وزن معکوس به آن اختصاص داده می‌شود، بدین صورت که برای کلاس با تعداد نمونه کمتر (مانند مرحله ۱ خواب) وزن بزرگتری اختصاص می‌یابد و بر عکس. بر این اساس فرمول نهایی تابع زیان آنتروپی متقاطع دسته‌ای، بصورت زیر تغییر می‌کند که تابع وزن دهی در آن نیز آورده شده است:

$$L = -\frac{1}{M} \sum_{i=1}^M \sum_{k=1}^K w_k \cdot y_{i,k} \cdot \log(p_{i,k}) \quad (4)$$

$$w_k = \frac{N}{C \cdot N_k}$$

(batch_size, T_frames, F_bins) است. در گام بعد، برای آن که مکانیزم توجه چندسر بتواند روابط میان باندهای فرکانسی را بیاموزد، محورهای زمان و فرکانس جابه‌جا می‌شوند تا محور فرکانس در موقعیت اصلی توجه قرار گیرد. پس از این بازآرایی، لایه توجه چندسر بر داده‌ها اعمال می‌شود تا مدل بتواند الگوها و وابستگی‌های پیچیده میان باندهای مختلف مانند دلتا، تتا، آلفا و بتا را شناسایی کند. سپس خروجی مکانیزم توجه برای بازگشت به ساختار اصلی خود، یعنی (batch_size, T_frames, F_bins)، مجدد بازآرایی می‌شود تا خروجی لایه توجه با ورودی اولیه به کمک یک اتصال باقیمانده^۱ جمع گردد و سپس نرمال‌سازی لایه‌ای^۲ بر نتیجه اعمال می‌گردد تا یادگیری مدل پایدارتر شود. در پایان، جهت کاهش ابعاد و فشردگی ویژگی‌های استخراج‌شده، از لایه‌ی GAP در امتداد محور زمانی استفاده می‌شود. این لایه، میانگین انرژی هر باند فرکانسی را در طول کل بازه‌ی زمانی محاسبه کرده و یک بردار فشرده با ابعاد (batch_size, F_bins) تولید می‌کند که نمایانگر اهمیت نسبی باندهای مختلف در سیگنال EEG است. این طراحی، برخلاف لایه‌های پیچشی که تنها ویژگی‌های محلی را استخراج می‌کنند، به مدل امکان می‌دهد روابط کلی میان باندهای فرکانسی (مانند هماهنگی بین باندهای دلتا و آلفا) را نیز بیاموزد و ویژگی‌هایی سطح بالا و قابل تفسیر از سیگنال EEG به دست آورد.

۲.۴ لایه‌های نهایی معماری شبکه

در قسمت انتهایی مدل، بردار ویژگی حاصل از مسیرهای زمانی و فرکانسی به یک لایه کاملاً متصل با ۱۲۸ نورون منتقل می‌شود. این لایه با هدف فشردگی و ترکیب مؤثر ویژگی‌های استخراج‌شده، آن‌ها را به ورودی مناسب‌تر برای فرآیند تصمیم‌گیری نهایی تبدیل می‌نماید. به منظور ارتقاء پایداری فرآیند آموزش و کاهش حساسیت مدل نسبت به مقادیر اولیه وزن‌ها، بلافاصله پس از این لایه از نرمال‌سازی دسته‌ای استفاده شده است. در ادامه، تابع فعال‌ساز غیرخطی GELU به کار گرفته می‌شود که با ارائه رفتاری نرم‌تر نسبت به ReLU، تعادل

جدول ۱. مشخصات داده‌های استفاده شده در این پژوهش (هر نمونه یک ایپاک ۳۰ ثانیه‌ای است)

Dataset	subjects	EEG Channel	Sampling rate	W	N1	N2	N3	REM	Total samples
Sleep_Edf_20	۲۰	Fpz-cz	100 HZ	۸۲۸۵ ٪۱۹.۶	۲۸۰۴ ٪۶.۶	۱۷۷۹۹ ٪۴۲.۱	۵۷۰۳ ٪۱۳.۵	۷۷۱۷ ٪۱۸.۲	۴۲۳۰۸
Sleep_Edf_78	۷۸	Fpz-cz	100 HZ	۶۵۹۵۱ ٪۳۳.۷	۲۱۵۲۲ ٪۱۱.۰۱	۶۹۱۳۲ ٪۳۵.۴	۱۳۰۳۹ ٪۶.۷	۲۵۸۳۵ ٪۱۳.۷۴	۱۹۵۴۷۹

² Layer Normalization

¹ Residual Connection

کانال Fpz-Cz و Pz-Oz با نرخ نمونه‌برداری ۱۰۰ هرتز، یک کانال EOG و یک کانال EMG چانه است. همراه با مطالعات قبلی [۱۷]، [۳۱]، [۳۲]، [۳۳]، [۳۴]، [۳۵]، داده‌های SC* انتخاب شدند و از کانال Fpz-Cz به عنوان ورودی مدل‌های پیشنهادی استفاده گردید.

در پیش پردازش داده‌ها ابتدا قسمتهای با برچسب (UNKNOWN) که به هیچ یک از مراحل خواب تعلق ندارد، حذف شدند. سپس قسمت‌های با برچسب مرحله ۳ و ۴ خواب در یک مرحله یعنی خواب عمیق یا N3 مطابق با استاندارد AASM ادغام شدند. لازم به ذکر است که فقط ۳۰ دقیقه از دوره‌های بیداری قبل و بعد از دوره‌های خواب مورد استفاده قرار گرفتند [۱۴]. برای ارزیابی عملکرد مدل‌های مختلف در طبقه‌بندی مراحل خواب، از چهار معیار اصلی شامل صحت (ACC)، امتیاز F1 میانگین کلان (MF1)، ضریب کاپای کوهن (κ) [۳۶] و میانگین هندسی کلان (MGm) استفاده گردید. دو معیار MF1 و MGm به‌طور گسترده برای ارزیابی مدل‌ها در مجموعه‌داده‌های نامتوازن مانند داده‌های مراحل خواب به کار می‌روند، زیرا عملکرد بهتری در چنین شرایطی نسبت به معیارهای کلاسیک دارند [۳۷-۳۸]. با توجه به مثبت‌های واقعی (TPi)، مثبت‌های کاذب (FPi)، منفی‌های واقعی (TNi) و منفی‌های کاذب (FNi) برای کلاس i ام، ACC، MF1 و MGm به صورت زیر تعریف می‌شوند:

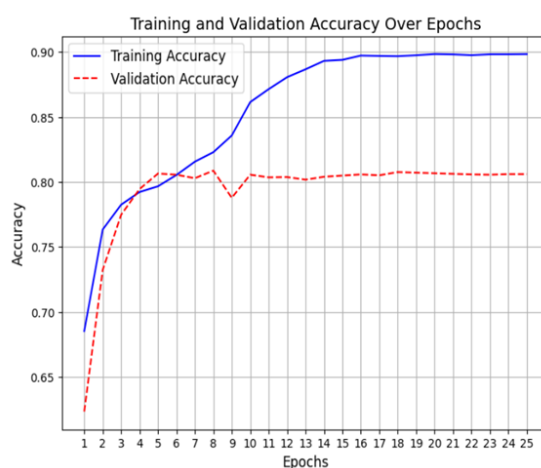
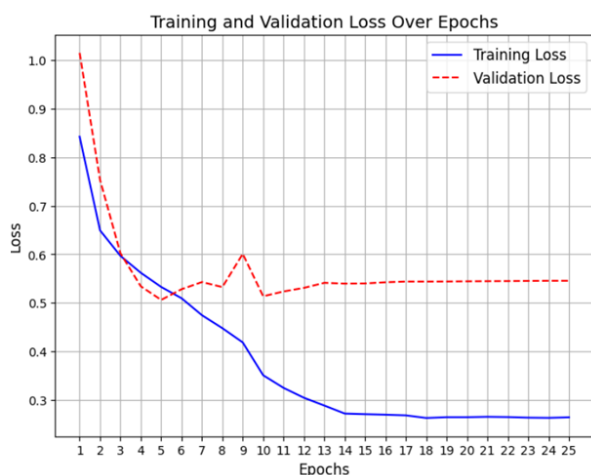
در رابطه فوق، N تعداد نمونه‌های کل کلاس‌ها، C تعداد کلاس‌ها و N_k تعداد نمونه‌های موجود در کلاس K است. استفاده از این تابع وزن دهی تاثیر زیادی در بهبود دقت‌ها به ویژه دقت مرحله ی یک خواب داشته است.

۳. پیاده سازی و نتایج

در این بخش به معرفی مجموعه داده‌های استفاده شده در این پژوهش، پیاده سازی مدل پیشنهادی و اعمال تغییراتی در آن و در پایان تحلیل نتایج بدست آمده پرداخته شده است.

۳.۱ مجموعه داده‌ها و معیارهای ارزیابی

در این پژوهش، از دو مجموعه داده عمومی، یعنی Sleep-EDF-20 و Sleep-EDF-78 از سایت فیزیوت استفاده شده است که جزییات آن در جدول ۱ آورده شده است [۳۰]. یک کانال EEG از هر ثبت در اجرای مدل پیشنهادی مورد استفاده قرار گرفته است. مجموعه Sleep-EDF-20 شامل ثبت داده از ۲۰ فرد است در حالی که Sleep-EDF-78 یک نسخه گسترش یافته با ۷۸ ثبت است. شرکت کنندگان در دو مطالعه شرکت داشتند. اولی SC*^۱ است که اثرات سن را بر خواب مطالعه می‌کند و روی شرکت کنندگان سالم با سنین ۲۵ تا ۱۰۱ سال انجام شد. دومی ST*^۲ است که اثرات تمازپام را بر خواب در ۲۲ مرد و زن قفقازی بدون مصرف هیچ داروی دیگری بررسی کرده است. برای این دو مجموعه داده، هر فایل PSG شامل دو



شکل ۳. مقایسه‌ی دقت و خطای آموزش و آزمون در مجموعه‌داده

$$Sensitivity = Recall_i = RE = \frac{TP_i}{TP_i + FN_i} \quad (۸)$$

$$Precision_i = PR = \frac{TP_i}{TP_i + FP_i} \quad (۹)$$

$$Specificity_i = SP = \frac{TN_i}{TN_i + FP_i} \quad (۱۰)$$

$$ACC = \frac{\sum_{i=1}^K TP_i}{M} \quad (۵)$$

$$MF1 = \frac{1}{K} \sum_{i=1}^K \left(\frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i} \right) \quad (۶)$$

$$MGm = \frac{1}{K} \sum_{i=1}^K \sqrt{Recall_i \times Specificity_i} \quad (۷)$$

² Sleep Telemetry

¹ Sleep Cassette



در این روابط، M نشان‌دهنده تعداد کل نمونه‌ها و K تعداد کلاس‌ها یا مراحل خواب است. علاوه بر این، دقت (PR)، حساسیت (RE)، امتیاز $F1$ (MF1) و میانگین هندسی (GM) هر کلاس نیز برای ارزیابی مدل پیشنهادی استفاده شد. این معیارها مشابه طبقه‌بندی دودویی محاسبه می‌شوند، به‌طوری که یک کلاس به‌عنوان کلاس مثبت و چهار کلاس دیگر به‌عنوان کلاس منفی در نظر گرفته می‌شوند [۱۷].

۳.۲ نتایج آزمایش‌ها

در انجام این پژوهش ۷۰ درصد از کل داده‌ها به‌عنوان داده‌های آموزش و ۳۰ درصد نیز به‌عنوان داده‌ی آزمون انتخاب شدند. همچنین ۱۵ درصد از داده‌های آموزش به‌اعتبارسنجی مدل‌ها اختصاص یافت. حداکثر تعداد دوره‌های آموزشی برای تمامی مدل‌های مورد آزمایش ۲۵ در نظر گرفته شد. نکته قابل توجه در استفاده از مکانیزم توجه در این پژوهش این است که با سعی و خطا مشاهده شد که مدل‌های مبتنی بر مکانیزم توجه معمولاً بعد از ۱۵ دوره آموزشی، به پایداری می‌رسند. این در حالی است که مدل LSTM برای رسیدن به وضعیت پایدار به حدود ۲۰ تا ۲۵ دوره آموزشی نیاز دارند. بنابراین، مکانیزم‌های توجه از نظر سرعت آموزش، نسبت به LSTM عملکرد بهتری دارند و بهینه‌تر عمل می‌کنند. برای پیاده‌سازی مکانیزم توجه در مدل پیشنهادی، از لایه‌ی MHA موجود در کتابخانه keras استفاده شده است. این لایه به‌صورت پیش‌فرض دارای لایه Dense است که ابعاد خروجی مکانیزم توجه را با ابعاد ورودی

توکن‌ها تطبیق می‌دهد. به همین دلیل، حتی با تغییر ابعاد توکن‌ها و تعداد سرها، سازگاری ابعادی حفظ می‌شود. همچنین به‌منظور افزایش تعمیم‌پذیری و کاهش خطر بیش‌برازش، در این بخش از Dropout با نرخ ۰.۱ استفاده شده است که سبب می‌شود مدل وابستگی بیش‌ازحد به داده‌های آموزشی پیدا نکند و عملکرد بهتری روی داده‌های آزمون داشته باشد. شکل ۳ روند یادگیری مدل پیشنهادی را نشان می‌دهد که شامل تغییرات تلفات (Loss) و دقت (Accuracy) در هر دور آموزش است. همانطور که مشاهده می‌شود، مدل از همگرایی سریع و پایداری برخوردار است. نکته قابل توجه، روند کاهشی پایدار تلفات آموزش در کنار تثبیت تلفات اعتبارسنجی است که نشان‌دهنده توانایی مدل در تعمیم‌پذیری مناسب و جلوگیری از بیش‌برازش^۱ است. این ویژگی به ویژه در کاربردهای تحلیل سیگنال‌های خواب که با داده‌های پیچیده سروکار دارند، حائز اهمیت است. این رفتار مطلوب مدل را می‌توان به ترکیب بهینه مکانیزم توجه و معماری طراحی شده نسبت داد.

کلیه آزمایش‌های این پژوهش در محیط محاسباتی Kaggle و با بهره‌گیری از واحد پردازش گرافیکی NVIDIA (GPU) P100 انجام شد. اندازه دسته ۱۲۸ در نظر گرفته شد و از بهینه‌ساز Adam استفاده گردید و به منظور بهبود همگرایی مدل و جلوگیری از توقف زود هنگام در مینیمم‌های محلی، در فرآیند آموزش از سازوکار کاهش نرخ یادگیری بر اساس پارامتر ReduceLROnPlateau استفاده شد. این روش به‌طور

جدول ۲. نتایج اعمال روشهای پیشنهادی روی مجموعه داده‌های Sleep-edf20.

دقت مدل‌های پیشنهادی در چهار حالت بررسی شد. حالت اول فقط شامل لایه‌های پیچشی به صورت مشخص شده در شکل ۲ است. در حالت دوم مکانیزم توجه چندسر زمانی، در امتداد بلوک CNN اضافه شد. در حالت سوم شاخه‌ی توجه فرکانسی، موازی با بلوک CNN در حالت اول اضافه گردید. حالت چهارم ترکیب حالت دوم و سوم است که در آن همزمان از مکانیزم توجه در امتداد لایه‌های CNN و همچنین توجه فرکانسی روی طیف زمان-فرکانس حاصل از STFT به موازات آن استفاده شده است. نتایج پیاده‌سازی حالت‌های مختلف در جدول ۲ آورده شده است. نتایج نشان می‌دهد مدل‌های ارائه شده در حالت سوم و چهارم بالاترین دقت را دارند از اینرو این دو مدل، روی دیتاست Sleep edf78 نیز بررسی شدند که نتایج آن در

خودکار نرخ یادگیری فعلی را در صورت عدم کاهش تابع هزینه اعتبارسنجی (val_loss) در طی ۴ دوره متوالی (patience=4)، در ضریب ۰.۱ (factor=0.1) ضرب کرده و به مقدار جدیدی کاهش می‌دهد. این فرآیند به مدل امکان می‌دهد تا در مراحل پایانی آموزش، با گام‌های کوچک‌تری پارامترهای خود را بهینه‌سازی کند و دقت نهایی افزایش یابد. نرخ یادگیری از دوره آموزشی دوازدهم با مقدار $1e-4$ شروع به کاهش نمود و در دوره آموزشی بیست و پنج به حداقل مقدار $1e-6$ رسید. این روش باعث افزایش پایداری آموزش و بهبود همگرایی مدل در مراحل پایانی گردید به‌ویژه در شرایطی که مدل به یک حالت شبه پایدار رسیده و نیاز به تنظیم دقیق‌تری در به روزرسانی وزن‌ها داشت.

¹ Overfitting

نشان‌دهنده تعداد نمونه‌های دسته بندی‌شده توسط متخصص خواب و هر ستون تعداد پیش‌بینی‌های مدل است. جداول همچنین مقادیر کلاس به کلاس برای دقت، RE، F1 و GM را نشان می‌دهند. نکته قابل توجه اینکه خواب مرحله ۱ پایین‌ترین عملکرد را دارد (F1) کمتر از ۵۰٪، زیرا از یکسو نمونه کمتری دارد و از سوی دیگر به دلیل شباهت سیگنال EEG در این مرحله با بیداری و خواب REM، به اشتباه دسته بندی می‌شود. در مقابل، مرحله ۲ خواب با بیشترین دقت در Sleep-EDF-20 دسته بندی گردید.

جدول ۳ آورده شده است. بر اساس نتایج بدست آمده، مدل اول در جدول ۳ (حالت سوم در جدول ۲) به دلیل سادگی و سبک بودن و همچنین دقت بهتر برای انجام آنالیزهای بیشتر انتخاب شد.

۳.۳ نتایج مدل پیشنهادی

جداول ۴ و ۵ ماتریس‌های آشفتگی^۱ مدل پیشنهادی را نشان می‌دهند که بر روی کانال Fpz-Cz در هر دو مجموعه داده Sleep-EDF اعمال شده است این ماتریس‌ها از مجموع امتیازهای آزمون، در ۲۵ دوره آموزشی حاصل شده‌اند. هر ردیف

جدول ۴. ماتریس آشفتگی مدل پیشنهادی اعمال شده بر کانال FPZ-CZ از مجموعه داده Sleep_Edf_20

Predicted						Per_class metrics			
	W	N1	N2	N3	REM	PR	RE	F1	GM
W	۱۹۱۷	۲۷۷	۴۹	۷۳	۱۷۰	۰.۹۵	۰.۷۷	۰.۸۵	۰.۸۷
N1	۴۱	۳۷۸	۱۱۱	۷	۳۲۳	۰.۴۲	۰.۴۵	۰.۴۳	۰.۶۵
N2	۲۵	۱۰۹	۴۴۲۱	۴۰۶	۳۷۹	۰.۹۱	۰.۸۳	۰.۸۷	۰.۸۸
N3	۵	۱	۵۸	۱۶۴۵	۲	۰.۷۷	۰.۹۶	۰.۸۶	۰.۹۵
REM	۳۰	۱۴۵	۲۰۰	۳	۱۹۳۷	۰.۶۹	۰.۸۴	۰.۷۶	۰.۸۷

جدول ۵. ماتریس آشفتگی مدل پیشنهادی اعمال شده بر کانال FPZ-CZ از مجموعه داده Sleep_Edf_78

Predicted						Per_class metrics			
	W	N1	N2	N3	REM	PR	RE	F1	GM
W	۱۶۲۶۴	۲۳۲۵	۱۸۴	۱۱۰	۹۰۲	۰.۹۶	۰.۸۲	۰.۸۸	۰.۹۰
N1	۳۹۸	۳۰۷۹	۱۲۴۷	۴۴	۱۶۸۹	۰.۳۷	۰.۴۷	۰.۴۲	۰.۶۵
N2	۸۴	۱۸۵۶	۱۵۶۶۳	۱۲۰۲	۱۹۳۵	۰.۸۷	۰.۷۵	۰.۸۰	۰.۸۴
N3	۱۰	۱۹	۴۱۶	۳۴۳۹	۲۸	۰.۷۱	۰.۸۸	۰.۷۹	۰.۹۲
REM	۸۱	۹۲۲	۵۰۸	۱۴	۶۲۲۵	۰.۵۷	۰.۸۰	۰.۶۷	۰.۸۵

سبک در تشخیص خودکار مراحل خواب باشد. با توجه به نتایج جدول ۶، مدل پیشنهادی SleepTFANet با وجود سادگی ساختار، عملکردی قابل رقابت با مدل‌های پیچیده‌تری نظیر AttnSleep، SleepEEGNet و LWSleepNet دارد. در مجموعه داده Sleep-EDF-20، مدل در تفکیک مرحله یک و دو خواب (N₁ و N₂) عملکرد قابل توجهی داشته و در معیارهای کلی (ACC=81.1%) و (MF1=75.3%) نتایجی نزدیک به مدل‌های مرجع حاصل شده است. در حالی که مدل‌هایی مانند LightSleepNet با وجود دقت کلی مناسب، در تشخیص مرحله یک خواب ضعف قابل ملاحظه‌ای دارند. در مجموعه داده Sleep-EDF-78 نیز مدل پیشنهادی در تشخیص بیداری (W) و مرحله دو خواب عملکرد مناسبی داشته است. هرچند دقت کلی آن (ACC=76.1%) اندکی کمتر از مدل‌های LWSleepNet و

۴. بحث و تحلیل نتایج

مدل پیشنهادی SleepTFANet با بهره‌گیری همزمان از مسیرهای زمانی متشکل از لایه‌های پیچشی، و توجه فرکانسی که حاصل اعمال مکانیزم توجه چند سر روی محتوای طیفی است بدون اتکا به ساختارهای سنگین محاسباتی، موفق به دستیابی به دقتی قابل قبول و رقابتی در طبقه‌بندی مراحل خواب شده است. این مدل با بهره‌برداری از ویژگی‌های ترکیبی و ادغام اطلاعات در دو حوزه زمان و فرکانس، تعادلی مناسب میان کارایی محاسباتی، دقت طبقه‌بندی و سادگی معماری ایجاد کرده است. به‌ویژه در شرایطی که منابع سخت‌افزاری محدود هستند یا هدف، استفاده از مدل در دستگاه‌های پوشیدنی باشد، SleepTFANet می‌تواند گزینه‌ای مناسب و

¹ Confusion matrix

نتایج نشان می‌دهد که SleepTFANet با وجود ساختار سبک‌تر و تعداد پارامتر کمتر نسبت به مدل‌های پیچیده‌تر، قادر است دقتی نزدیک یا حتی رقابتی در هر دو مجموعه داده Sleep-EDF-20 و Sleep-EDF-78 ارائه دهد.

AttnSleep بوده، اما در شاخص MGm با مقدار ۸۳.۶٪ بالاترین عملکرد را میان تمام مدل‌ها دارد. این موضوع نشان‌دهنده تعادل مناسب مدل در شناسایی مراحل مختلف خواب و توانایی آن در حفظ دقت میان کلاس‌های نامتوازن است. به‌طور کلی،

جدول ۶

Dataset	Method	Per_Class F1_scoer					Overall metetics			
		W	N1	N2	N3	REM	ACC	MF1	K	MGm
Sleep_Edf_20	LightSleepNet[16]	۹۰	۳۱.۰	۸۸.۰	۸۹.۰	۷۸.۰	۸۳.۸	۷۵.۳	۰.۷۸	گزارش نشده
	SleepEEGNet[18]	۸۹.۴	۴۴.۴	۸۴.۴	۸۴.۶	۷۹.۶	۸۱.۵	۷۶.۶	۰.۷۵	۸۵.۳
	AttnSleep[17]	۸۹.۷	۴۲.۶	۸۸.۸	۹۰.۲	۷۹.۰	۸۴.۴	۷۸.۱	۰.۷۹	۸۵.۵
	LWSleepNet[15]	۹۲.۴	۴۱.۳	۹۰.۲	۸۸.۴	۸۴.۰	۸۶.۶	۷۹.۲	۰.۸۱	گزارش نشده
	SleepTFNet (Ours)	۸۶.۴	۴۱.۷	۸۷.۰	۸۶.۲	۷۵.۴	۸۱.۱	۷۵.۳	۰.۷۵	۸۴.۹
Sleep_Edf_78	SleepEEGNet[18]	۸۹.۸	۴۲.۱	۷۵.۲	۷۰.۴	۷۰.۶	۷۴.۲	۶۹.۶	۰.۶۶	۸۲.۳
	AttnSleep[17]	۹۲.۰	۴۲.۰	۸۵.۰	۸۲.۱	۷۴.۲	۸۱.۳	۷۵.۱	۰.۷۴	۸۳.۶
	LWSleepNet[15]	۹۲.۲	۴۰.۲	۸۶.۹	۷۸.۴	۷۳.۸	۸۱.۵	۷۴.۳	۰.۷۵	گزارش نشده
	SleepTFNet (Ours)	۸۸.۸	۴۲.۰	۸۰.۰	۷۸.۸	۶۷.۱	۷۶.۱	۷۱.۵	۰.۶۸	۸۳.۶

جدول ۷

Model	SleepTFNet (Ours)	LWSleepNet	SleepEEGNet	LightSleepNet	TinySleepNet [12]	AttnSleep
Param	2.2×10^4	1.8×10^5	2.1×10^7	4.3×10^4	1.3×10^6	5.2×10^5

توسط شبکه عصبی پیچشی یک بعدی و اطلاعات حاصل از اعمال مکانیزم توجه چند سر روی اسپکتروگرام سیگنال EEG، برای طبقه بندی مراحل خواب استفاده شد. تک فرکانس‌هایی که با زمان تغییر می‌کنند به عنوان توکن‌ها در مکانیزم توجه مورد استفاده قرار گرفتند و تعداد و نوع فیلترها در شبکه عصبی پیچشی به گونه ای تعیین شدند که نتایج بهبود یابد. در نهایت از ترکیب اطلاعات حاصل از این فرآیند با نقشه های ویژگی استخراج شده از شبکه های عصبی پیچشی یک بعدی، نتایج مطلوبی در تشخیص مراحل خواب حاصل شد. بعلاوه به‌منظور مقابله با مشکل عدم توازن داده‌ها در مراحل مختلف خواب، در فرآیند آموزش از وزن‌دهی کلاس‌ها استفاده شده است. این روش با تخصیص وزن‌های بالاتر به کلاس‌های کم‌نمونه (مانند مرحله‌ی یک)، تأثیر نابرابری توزیع داده‌ها را در تابع زیان، کاهش می‌دهد. نتایج مطلوب این مدل می‌تواند به دلیل توجه به تغییرات فرکانسی سیگنال EEG در زمانهای مختلف باشد مشخصه‌ای که می‌تواند در شناسایی خودکار مراحل خواب از

طبق جدول ۷ مدل پیشنهادی، با تنها ۲ هزار پارامتر یکی از سبک‌ترین معماری‌ها در حوزه تشخیص مراحل خواب است. در مقایسه با مدل‌هایی نظیر SleepEEGNet با ۲۱ میلیون پارامتر، TinySleepNet با ۱.۳ میلیون پارامتر و حتی مدل‌های سبک‌تری مانند LightSleepNet با ۴۳ هزار پارامتر و LWSleepNet با ۱۸۰ هزار پارامتر، SleepTFANet به‌طور قابل توجهی ساده‌تر و بهینه‌تر است. این کاهش چشمگیر در تعداد پارامترها نه تنها منجر به کاهش مصرف حافظه و انرژی می‌شود، بلکه امکان استفاده در دستگاه‌های قابل حمل و سیستم‌های زمان حقیقی را نیز فراهم می‌سازد. همچنین، سبک بودن مدل موجب افزایش سرعت آموزش و پیش‌بینی شده و آن را به گزینه‌ای مناسب برای کاربردهای بالینی و پایش لحظه‌ای خواب تبدیل کرده است.

روند متداول استفاده از مکانیزم توجه، اعمال آن روی سری زمانی یا نقشه های ویژگی حاصل از شبکه های پیچشی است. در این پژوهش برای اولین بار از ادغام اطلاعات استخراج شده

منابع

1. Carskadon, M.A., & Dement, W.C. (2011). Monitoring and staging human sleep. In M.H. Kryger, T. Roth, & W.C. Dement (Eds.), *Principles and practice of sleep medicine*, 5th edition, (pp 16-26). St. Louis: Elsevier Saunders.
2. Jahrami HA, Alhaj OA, Humood AM, et al (2022) Sleep disturbances during the covid-19 pandemic: systematic review, meta-analysis, and meta-regression. *Sleep Medicine Reviews* 62:10159
3. Zhang Y, Ren R, Yang L, et al (2022b) Sleep in alzheimer's disease: a systematic review and meta analysis of polysomnographic findings. *Translational Psychiatry* 12(1):136
4. Baglioni C, Battagliese G, Feige B, et al (2011) Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. *Journal of Affective Disorders* 135(1-3):10-19
5. Clarke G, Harvey AG (2012) The complex role of sleep in adolescent depression. *Child and Adolescent Psychiatric Clinics* 21(2):385-400.
6. مارکوس م، آزاد صفت سن، عراقیان س اختلالات خواب: ارزیابی، تشخیص و درمان در یک نگاه. تهران: انتشارات ارجمند؛ ۱۴۰۰
7. Bloch, K.E.: Polysomnography: a systematic review. *Technology and Health Care* (4), 285-305 (1997)
8. Wulff K, Gatti S, Sleep and Wettstein JG, et al (2010) circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nature Reviews Neuroscience* 11(8):589-599
9. Kayabekir M (2019) Sleep physiology and polysomnogram, physiopathology and symptomatology in sleep medicine. In: *Updates in Sleep Neurology and Obstructive Sleep Apnea*. IntechOpen
10. Khalighi, S., Sousa, T., Santos, J.M., Nunes, U.: Isruc-sleep: A comprehensive public dataset for sleep researchers. *Computer Methods and Programs in Biomedicine*, 180-192 (2016)
11. Zhang R, Tian D, Xu D, et al (2022a) A survey of wound image analysis using deep learning: classification, detection, and segmentation. *IEEE Access* 10:79502-79515
12. Supratak A, Dong H, Wu C, Guo Y. TinySleepNet: An efficient deep learning model for sleep stage scoring with single-channel EEG. *Comput Biol Med*. 2020;124:103918.
13. Korkalainen H, Saeed S, Himanen SL, et al. Accurate sleep stage classification with machine learning using EEG and respiratory signals. *Sleep Breath*. 2020;24(4):1285-1295.
14. Supratak A, Dong H, Wu C, Guo Y. DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Trans Neural Syst Rehabil Eng*. 2017;25(11):1998-2008.
15. Yang X, Zhang Y, Liang Y, et al. LWSleepNet: Lightweight multiresolution CNN with attention for sleep stage classification. *IEEE Trans Instrum Meas*. 2023;72:1-10.
16. Liao Y, Liang S, Wang Y, et al. LightSleepNet: A lightweight deep learning model for sleep stage classification. *Biomed Signal Process Control*. 2021;68:102739.
17. Eldele E, Chen Z, Liu C, et al. An attention-based deep learning approach for sleep stage classification with single-channel EEG. *IEEE Trans Neural Syst Rehabil Eng*. 2021;29:809-818.
18. Mousavi S, Afghah F, Acharya UR. SleepEEGNet: Automated sleep stage scoring with sequence-to-sequence deep learning approach. *PLoS One*. 2019;14(5):e0216456.
19. Ji D, Wang Y, Liu M, et al. MixSleepNet: Multi-modal fusion of EEG, EOG, EMG and ECG for sleep stage classification using GCN and 3D-CNN. *IEEE J Biomed Health Inform*. 2024;28(1):56-67.
20. Li Y, Zheng H, Wang W, et al. A novel deep learning model for imbalanced sleep stage classification: 1D-ResNet-SE-LSTM with class-weighted loss. *Expert Syst Appl*. 2023;212:118715.
21. Zhou J, Xu T, Chen X, et al. GAN-enhanced sleep staging with white Gaussian noise and CNN-BiLSTM for class imbalance problem. *Comput Methods Programs Biomed*. 2022;217:106756.
22. Khalili S, Farokhniaee A, Asl BM. A novel hybrid CNN-TCNN framework for sleep stage classification. *Comput Methods Programs Biomed*. 2021;208:106243.
23. Chen X, Yu J, Zhang Z, Guo X, Peng H, Wang L. MSA-CNN: A lightweight multi-scale attention CNN model for sleep stage classification. *Biomed Signal Process Control*. 2024;86:105552. doi:10.1016/j.bspc.2023.105552.

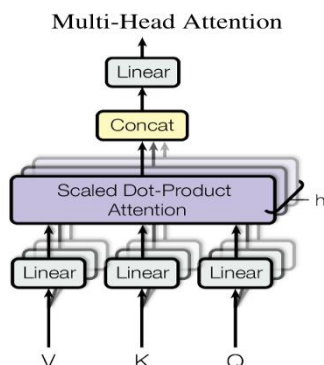
سیگنال EEG بسیار موثر باشد. زیرا در مراحل مختلف خواب، الگوهای فرکانسی متنوعی در زمانهای مختلف در سیگنال EEG آشکار می‌شوند. به عنوان مثال دوک های خواب در مرحله ۲ خواب یا امواج اهسته دلتا که با دامنه زیاد در خواب عمیق غالب هستند. عملکرد قابل قبول مدل پیشنهادی SleepTFANet، با پیاده سازی چند ترکیب متفاوت از استفاده همزمان مکانیزم توجه و شبکه عصبی پیچشی نشان داده شده است.

۵. خلاصه و نتیجه گیری

در این پژوهش اطلاعات از سیگنال EEG در دو مسیر موازی استخراج شدند. در یک مسیر با انتخاب مناسب اندازه فیلترها در لایه های پیچشی تلاش شد اطلاعات به نحو مناسبی استخراج گردد و در مسیر دیگر با اعمال مکانیزم توجه در فرکانسهای مختلف روی اسپکتروگرام حاصل از STFT، ویژگی های مکمل استخراج و مورد استفاده قرار گرفتند. این مدل ترکیبی موازی، با تعداد پارامتر کمتر نسبت به مدل های ارائه شده در ادبیات، توانست نتایج خوبی در طبقه بندی خودکار مراحل خواب ارائه دهد. در مجموع، یافته های این تحقیق نشان داد که ترکیب همزمان ویژگی های زمانی و فرکانسی به همراه توجه در محور فرکانس، می تواند به بهبود قابل توجه عملکرد مدل های طبقه بندی مراحل خواب منجر شود. اعمال مکانیزم توجه به خود چند سر در فرکانسهای مختلف از اسپکتروگرام حاصل از STFT برای اولین بار، در این پژوهش بررسی شد. استفاده همزمان از مکانیزم توجه و STFT می تواند بیشتر بررسی شود به گونه که توجه به قسمتهای حقیقی و موهومی یا به اندازه و فاز سیگنال بعد از اعمال STFT، بررسی گردد. در مطالعات آینده می توان از روشهایی مانند Grad Cam برای استخراج نقشه های حساسیت استفاده نمود تا بتوان به صورت شهودی از طریق استخراج نقشه های حساسیت، محل هایی در سیگنال EEG را که توسط مکانیزم توجه برجسته شده است آشکار نمود. بعلاوه از آنجایی که عدم توازن تعداد نمونه ها در کلاسها مختلف می تواند روی نتایج تاثیر نامطلوب داشته باشد در مطالعات آینده می توان از توابع هزینه دیگر مانند هزینه کانونی^۱ استفاده نمود. همچنین تکنیک های تولید داده مانند خودرمزگذار متغیر و شبکه های مولد متخاصم می تواند برای متوازن سازی مورد استفاده قرار گیرد. سرانجام از مدل ارائه شده در این پژوهش به دلیل داشتن پارامتر کمتر می توان در دستگاه های پوشیدنی و سیستم های پایش خواب خانگی استفاده نمود.

¹ Focal loss

ویژگی‌ها است. از X با سه نداشت خطی، ورودی مناسب جهت استفاده در بلوک توجه به صورت زیر ایجاد می‌شود. (در مدل پیشنهادی ورودی به گونه ای مورد استفاده قرار می‌گیرد که از طیف زمان فرکانس، توکن‌ها در محور فرکانس انتخاب می‌شوند یعنی تک فرکانسهایی که با زمان تغییر می‌کند).



شکل ۳: ساختار مکانیزم توجه چندسر [۲۷]

$$Q = W_Q X, K = W_K X, V = W_V X \quad (1)$$

در این رابطه W_Q, W_K, W_V ماتریس‌های وزن با ابعاد مناسب هستند و $Q, K, V \in \mathbb{R}^{L \times d}$ و ماتریس‌های پرسش، کلید و ارزش هستند. برای هر سر، توجه به صورت زیر محاسبه می‌شود:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

در ساختار توجه چندسر، پارامتر d_k نمایانگر ابعاد بردارهای پرسش و کلید در هر سر توجه و تعیین‌کننده ظرفیت مدل در نمایش ویژگی‌های وابسته به زمان است. در نهایت خروجی تمامی سرهای توجه به هم متصل (Concat) شده و از طریق یک نداشت خطی نهایی (با پارامترهای W_O) به صورت زیر ترکیب می‌شوند:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_O \quad (3)$$

که در آن h تعداد سرهای توجه است [۲۷].

در MHA چهار لایه Dense وجود دارد، طیف حاصل از STFT به عنوان ورودی مکانیزم توجه چند سر است. برای اینکه توجه بتواند روابط بین باند‌های فرکانسی را بیاموزد ابتدا هر باند فرکانسی به بردار ویژگی قابل یادگیری (مشابه token embedding در مدل‌های زبانی) تبدیل شود که این عمل به وسیله سه لایه dense اول انجام می‌شود. بعد از محاسبه توجه در محور فرکانس، خروجی MHA ابعاد (batch, F, key_dim) دارد. اما برای بازگرداندن ساختار به محور زمانی (جهت ادغام با ویژگی‌های زمانی)، نیاز به یک نداشت است. لایه Dense نهایی در این ساختار، این نداشت (بازسازی ساختار زمان-فرکانس) را انجام می‌دهد.

اینکه طیف حاصل از STFT ترانهاده می‌شود تنها مرتبط به کتابخانه ای است که در این پژوهش استفاده شده است با ترانهاده کردن ماتریس STFT، هر توکن یک فرکانس موجود در طیف حاصل از STFT یا به عبارتی یک فرکانس از قطعه سیگنال EEG است که اطلاعات از آن استخراج می‌شود. توکن‌ها در ماتریس‌های وزن (W^Q و W^K و W^V) ضرب می‌شوند تا Token embedding بدست آید این فرآیند برای آموزش پذیر کردن فرآیند است. در ادامه شباهت یا همبستگی بین فرکانس مورد نظر با همه فرکانسها محاسبه و اعداد حاصل نرمالیزه می‌گردند در نهایت فرکانسها با این اعداد وزندهی شده و به خروجی می‌روند در پایان دوباره ماتریس STFT وزندهی شده، ترانهاده شده (تا به شکل اول خود بازگردد البته هر فرکانس در آن وزن ضرب شده است) و به خروجی می‌رود.

24. ابراهیمی ف، میکائیلی م، ناظریان ه- طبقه بندی خودکار مراحل خواب با استفاده از شبکه‌های عصبی مصنوعی. چهاردهمین کنفرانس مهندسی پزشکی ایران: ۱۳۸۶: تهران. شناسه ملی: ICBME14_085.

25. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaria J, Fadhel MA, Al-Amidie M, Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data*. 2021;8:53. doi:10.1186/s40537-021-00444-8.
26. J. B. Allen and L. R. Rabiner, "A unified approach to short-time Fourier analysis and synthesis," *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1558–1564, Nov. 1977, doi: 10.1109/PROC.1977.10770.
27. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2017;30:5998–6008.
28. Goodfellow I, Bengio Y, Courville A. *Deep learning*. Cambridge, MA: MIT Press; 2016.
29. King G, Zeng L. Logistic regression in rare events data. *Political Analysis*. 2001;9(2):137–63
30. A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.
31. O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, "Automatic sleep stage scoring with single-channel EEG using convolutional neural networks," 2016, *arXiv:1610.01683*. [Online]. Available: <http://arxiv.org/abs/1610.01683>
32. M. Sokolovsky, F. Guerrero, S. Paisarnrisomsuk, C. Ruiz, and S. A. Alvarez, "Deep learning for automated feature discovery and classification of sleep stages," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 6, pp. 1835–1845, Nov. 2020.
33. H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "Joint classification and prediction CNN framework for automatic sleep stage classification," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1285–1296, May 2019.
34. F. Li *et al.*, "End-to-end sleep staging using convolutional neural network in raw single-channel EEG," *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, Art. no. 102203
35. H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "SeqSleepNet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 400–410, Mar. 2019.
36. J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.
37. Y. Tang, Y.-Q. Zhang, N. V. Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 281–288, Feb. 2009.
38. He, H., & Garcia, E. A. (2009). *Learning from imbalanced data*. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284.

پایوست

مکانیزم توجه چند سر (MHA) از مدل ترنسفورمر [۲۷-۱۷] الهام گرفته شده است که به دلیل توانایی یادگیری روابط بلندمدت در جملات، موفقیت چشمگیری در کاربردهای ترجمه ماشینی نشان داده است [۱۷]. در مکانیزم توجه بردارهایی تحت عنوان توکن (پرسش (Query)، کلید (Key) و ارزش (Value)) تعریف می‌شوند (شکل ۳). پرسش نشان‌دهنده‌ی اطلاعاتی است که مدل به دنبال یافتن آن در سیگنال است. کلید توصیف ویژگی‌های هر قسمت از سیگنال است که با پرسش مقایسه می‌شود. ارزش، اطلاعات محتوایی آن قسمت از سیگنال است که در صورت مرتبط بودن، در خروجی مدل نقش دارد. با مقایسه پرسش و کلیدها، وزن توجه برای هر بخش محاسبه شده و با ارزش‌ها ترکیب می‌شود تا مدل بتواند تمرکز خود را بر قسمت‌های مهم‌تر سیگنال افزایش دهد. در این پژوهش از مکانیزم توجه به خود استفاده شده است که در آن شباهت یا ضرب نقطه ای بین هر توکن با همه توکن‌ها محاسبه شده و این مقادیر توسط تابع Softmax نرمالیزه می‌گردند و اینگونه اعدادی بدست می‌آیند که ارزش یا وزن هر توکن را در ایجاد بردار خروجی مشخص می‌کنند بدین ترتیب، ابتدا ورودی‌ها به صورت $X \in \mathbb{R}^{(L \times d)}$ بازآرایی شدند که در آن L طول توالی و d تعداد